

VISIT...

LANZAROTE  
*Caliente*.COM

# **Business Intelligence**

*The Ultimate Guide to BI, Artificial Intelligence,  
Machine Learning, Big Data, Cybersecurity, Data  
Science, and Predictive Analytics*

**© Copyright 2020**

All Rights Reserved. No part of this book may be reproduced in any form without permission in writing from the author. Reviewers may quote brief passages in reviews.

Disclaimer: No part of this publication may be reproduced or transmitted in any form or by any means, mechanical or electronic, including photocopying or recording, or by any information storage and retrieval system, or transmitted by email without permission in writing from the publisher.

While all attempts have been made to verify the information provided in this publication, neither the author nor the publisher assumes any responsibility for errors, omissions or contrary interpretations of the subject matter herein.

This book is for entertainment purposes only. The views expressed are those of the author alone, and should not be taken as expert instruction or commands. The reader is responsible for his or her own actions.

Adherence to all applicable laws and regulations, including international, federal, state and local laws governing professional licensing, business practices, advertising and all other aspects of doing business in the US, Canada, UK or any other jurisdiction is the sole responsibility of the purchaser or reader.

Neither the author nor the publisher assumes any responsibility or liability whatsoever on the behalf of the purchaser or reader of these materials. Any perceived slight of any individual or organization is purely unintentional.

# Table of Contents

## [Part 1: Business Intelligence](#)

[\*An Essential Beginner's Guide to BI, Big Data, Artificial Intelligence, Cybersecurity, Machine Learning, Data Science, Data Analytics, Data Mining, Social Media and Internet Marketing\*](#)

### [Introduction](#)

### [Chapter 1: An Introduction to BI, Big Data, AI, and More](#)

[Enter the Information Revolution](#)

[The World of Big Data Begins Taking Shape](#)

[What is Business Intelligence?](#)

[Artificial Intelligence and Machine Learning](#)

[Data Science](#)

[Data Mining](#)

[IoT](#)

[Data Analytics](#)

[Cybersecurity](#)

[Social Media](#)

[Internet Marketing](#)

### [Chapter 2 : An Overview of Business Intelligence](#)

[The Key Characteristics of BI for business](#)

[The Four-Step Process of Business Intelligence](#)

[Business Intelligence Applications](#)

[Reporting and Database Querying](#)

[Spreadsheets](#)

[Data Dashboards](#)

[OLAP](#)

[Data Mining](#)

[Data Warehouse](#)

[Data Cleansing](#)

[Business Activity Monitoring](#)

[Real-Time BI](#)

[The Core Benefits of Business Intelligence](#)

[Past or Future Data?](#)

### [Chapter 3: BI and AI](#)

[The core uses of AI](#)

[AI Can Make Decisions on Its Own](#)

[Process Automation](#)

[Cognitive Insights](#)

[Cognitive Engagement](#)

[How AI will Take BI to a new level](#)

#### [Chapter 4: BI and Big Data](#)

[Big Data Classifications](#)

[Structured Data](#)

[Unstructured Data](#)

[Semi-Structured Data](#)

[The Five V's of Big Data](#)

[How do Big Data and Business Intelligence Compare?](#)

[Big Data Resources](#)

[Developing a big data strategy](#)

[Problems with Big Data](#)

[Small Business Case Study: Advertising Using Lookalike Audiences](#)

#### [Chapter 5: BI And Machine Learning](#)

[What is Machine Learning?](#)

[Classifications of Machine Learning](#)

[Training Machine Learning Systems](#)

[Supervised Learning](#)

[Issues with Supervised Learning](#)

[Unsupervised Learning](#)

[Semi-Supervised Learning](#)

[Some Current Applications of Machine Learning](#)

#### [Chapter 6: Data Science](#)

[Data and Machines in Business](#)

[Comparing and Contrasting Data Science and Business Intelligence](#)

[Improved Efficiency Through Automation](#)

[Self-Service BI](#)

[Labor Saving](#)

#### [Chapter 7: Cybersecurity](#)

[Determining Infosec Needs of the Company](#)

[Cybersecurity in a nutshell](#)

[The Approach to Cybersecurity](#)

[Data Science and Cybersecurity](#)

[The Three Tenets of Cybersecurity](#)  
[Threat Detection](#)

[Chapter 8: BI and Data Mining](#)

[Data Mining](#)  
[Data Analytics](#)

[Chapter 9: BI and Social Media](#)

[Leveraging Social Media](#)  
[Web Scraping Tools](#)  
[Direct Interaction with Customers](#)  
[The Challenges of Social Media](#)  
[Data Issues with Social Media](#)  
[What Can You Get from Social Media?](#)  
[Dashboarding the Data](#)  
[Three Ways to Use Social Media](#)

[Chapter 10: BI and Internet Marketing](#)

[Social Media and Internet Platforms Best Suited for Online Marketing](#)  
[Customer Targeting](#)  
[Search Engine Marketing](#)  
[Establishing an Internet Presence](#)  
[Using Business Intelligence to Analyze and Improve](#)  
[Identify Weak Spots](#)  
[Is internet marketing for everyone?](#)

[Conclusion](#)

[Part 2: Predictive Analytics](#)

[\*The Secret to Predicting Future Events Using Big Data and Data Science\*](#)  
[\*Techniques Such as Data Mining, Predictive Modelling, Statistics, Data\*](#)  
[\*Analysis, and Machine Learning\*](#)

[Introduction](#)

[Chapter 1: The Basics of Predictive Analysis](#)

[Chapter 2: How to Predict Future Events with the Help of Big Data](#)

[Chapter 3: Can I Predict Future Events with Data Mining?](#)

[Chapter 4: Predicting Future Events with Predictive Modeling](#)

[Chapter 5: Using Statistics to Help Predict Future Events](#)

[Chapter 6: Bringing Data Analysis Into the Mix to Make Predictions](#)

[Chapter 7: Using Machine Learning to Make the Predictions For You](#)

[Chapter 8: How to Avoid Prediction Traps, Avoid Bias, and Make the Best](#)

[Predictions](#)

[Chapter 9: Top Reasons to Implement predictive analysis](#)

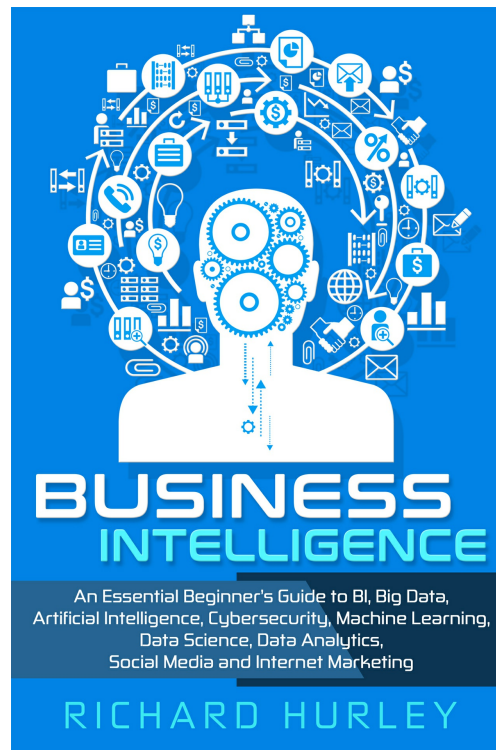
[Chapter 10: The Top Steps to Help You Create Your Own Predictive Analysis Model](#)

[Conclusion](#)



## Part 1: Business Intelligence

*An Essential Beginner's Guide to BI, Big Data,  
Artificial Intelligence, Cybersecurity, Machine  
Learning, Data Science, Data Analytics, Data Mining,  
Social Media and Internet Marketing*



# Introduction

Hello and welcome to *Business Intelligence - An Essential Beginner's Guide to BI, Big Data, Artificial Intelligence, Cybersecurity, Machine Learning, Data Science, Data Analytics, Data Mining, Social Media and Internet Marketing*.

In this book, we are going to introduce you to the concept of business intelligence. Business intelligence is closely connected to data, and so as part of our journey, we are going to see how modern companies are using big data, together with artificial intelligence, machine learning, and pattern recognition in order to build systems of data-driven decision making.

In the modern business world, the pace of action continues to quicken. Businesses need to be able to get actionable insights from their data in order to make the right decisions to act rapidly and effectively.

After we get a handle on what business intelligence is, we will learn how it ties together with the next computer revolution (happening right before our eyes). We will discuss the use of artificial intelligence, machine learning, and pattern recognition to retrieve insights from the large amounts of data that companies are routinely collecting, storing, and using.

We will also explore the roles that social media and internet marketing are playing in the growth of business intelligence, and how companies like Facebook are helping other companies act on data in powerful ways.

Previous technical revolutions were revolutions of *things*, and so they were easy to notice – and that revolution is still going on. Now it is being joined by a new revolution driven by data together with artificial intelligence and machine learning. While we are all touched by it, it's often not something we can see or touch! In that way, it is proceeding right under our noses, with many people barely aware of the massive and rapid changes taking place.

In this book, I hope to introduce you to these radical changes and get you excited about them. Business is in the process of being revolutionized in ways that it hasn't been since the industrial revolution.

# **Chapter 1: An Introduction to BI, Big Data, AI, and More**

In the 17th and 18th centuries, a merger of science, technology, and capitalism gave birth to the industrial revolution. Up to that time, societies had been static. Even during the times of the great ancient civilizations in Rome, Greece, Egypt, India, and China, societies made some strides but continued their existence in relatively static form. Poverty rates remained the same from the fall of the Roman empire until the start of the industrial revolution.

As soon as the first machines were invented, society began to take off in remarkable ways. Production of goods increased as mechanical devices enhanced the capabilities of workers to manufacture everything from textiles to pins. The efficiency of factories exploded overnight, so much so that by the middle of the 19th century, the world had been completely transformed. Even by then, people were already leading radically different lives than their grandparents had done, and the revolution was only getting started.

In this chapter, after a quick review, we will briefly introduce many of the main topics discussed in this book.

## **Enter the Information Revolution**

This revolution, as you well know, continued into the 20th century. The industrial part of the revolution continued growing, with the invention of the airplane, the automobile, the radio, television, and the distribution of electrical power into every home (at least in developed countries). But in the second world war, one of the most significant inventions of all time was quietly developed and used for some specialized tasks, such as breaking Nazi codes. This was an electronic computer. The genius engineer, Alan Turing, quickly recognized that the computer was an entirely different kind of machine. Its functions were sufficiently general that it could - at least, in theory - be imagined to work on any task. It was also something that processed information. All other machines developed to that point processed other things.

Turing recognized that a computer could mimic human intelligence. Whether or not it would be “real” intelligence was beside the point. This was a very powerful new machine, and after the war ended, the invention of the computer chip (and later, the microprocessor) ensured that computers were

here to stay and that they would develop very rapidly.

The development of computer power progressed, just like anything on an exponential curve does: slow at first and gradually picking up speed. In the early days, processors were much slower than they are now, but it was memory and the ability to store data that were really the weak points in computers until the 21st century. In the beginning, people were storing information on punch cards, and then feeding those into the computer. Later, this was replaced by magnetic tape, in the first transformation from physical to purely electronic form.

As the years went by, businesses were always collecting data. But in the old world before computers, the data was hard to collect. Even though some businesses were collecting some data, it was hard to do anything with it. The bottom line was that the data was impossible to bring together and synthesize, much less to interpret and gather any insights from.

This began changing when the world wide web (or internet) arrived. Likewise, computer storage capacity and the ability to retrieve and look at data began to improve. The internet only accelerated the pace of data collection. At first, people weren't even aware this was going on, and those collecting the data weren't sure what to do with it.

## **The World of Big Data Begins Taking Shape**

But the real story was the fact that computer memory was rapidly becoming very cheap and the capacity to store information was growing exponentially. This development was joined by advances in artificial intelligence, machine learning, and pattern recognition. The public had a fuzzy awareness of this, brought to light by occasional developments such as IBM computers defeating chess champions.

But the real action was in machine learning and the processing of data. Computer scientists had been strapped for years when it came to the development of artificial intelligence, but new paths forward came to light in the 1990s and later. Machine learning came to the forefront, and people began using computers in new ways. Rather than programming computers to do specific tasks, researchers were feeding them reams of data and letting the computers learn for themselves. It was found that computers were good at detecting patterns in data that people wouldn't be able to piece together because the data was too big - and apparently too random - for the human mind to comprehend or work with. Often the patterns that the machine

learning systems were able to detect in data led to actionable insights by companies. The massive amounts of data that were being collected by companies (as well as by governments and other institutions) became known as “big data”. Big data is more than just the data itself; this term incorporates the analysis of the data to reveal hidden patterns, trends, and relationships that exist. Typically, big data is related to the behaviors of people. You can think of it as the new gold of the 21st century.

This is where business intelligence began to come into full focus.

## **What is Business Intelligence?**

We associate business intelligence with computers, data, advanced statistics, and machine learning, but the reality is business intelligence has been with us ever since people have been in business. In the distant, and even more recent past, businesses had a limited capacity to collect data and take actions based on what the data was telling them.

*Business intelligence* is an incredibly broad category, and yet it plays a central role in modern business. Business intelligence begins with data collection and storage. Then it incorporates any activity, tool, or process that is used to extract actionable information and make decisions based on that information. Business intelligence can be anything from running reports on data that has been collected in the organization, to examining email communications between employees to develop better ways to foster collaboration, to using machine intelligence to improve the efficiency of logistics. The development of business intelligence has been propelled forward by advances in pattern recognition, machine learning, and artificial intelligence, along with the development of a new field that integrates many of these things under one umbrella called “data science”.

Some people have likened business intelligence and big data to the fourth industrial revolution – and with good reason. Big data, artificial intelligence, and machine learning are going to be shaping the future of business and of the structure of society at large in the future. Business intelligence is sometimes referred to as “BI,” following the shortened naming of artificial intelligence as “AI.”

## **Artificial Intelligence and Machine Learning**

Decades ago, Turing developed a simple test to discover if one could determine whether someone hidden behind a barrier was a human being or an

artificially intelligent computer. This gave birth to the idea of artificial intelligence, whereby computer systems would be built in ways that mimic the structure of the human brain, and they would learn the way a human brain does, rather than the simplistic ways that early computers did.

Developments began in fits and starts at first! In the late 1950s, an IBM engineer built a computer system that was able to learn how to play a game of checkers. His approach was completely different than the one normally used to program computers. Rather than program the computer with thousands of lines of computer code which ran by rules and told the program what to do step-by-step, he developed his checkers-playing software in a way that it would not be programmed at all. Rather, it would learn from experience. Put another way, it would learn by being exposed to data. The term ***machine learning*** was used to describe how the system worked.

This first attempt at machine learning was solving a simple problem, but it was remarkably successful. The more the machine played checkers, the better it got. It seemed that the adage “practice makes perfect” applied to this intelligent computer system as well.

Machine learning is a subfield, or subset, of the field of artificial intelligence. This is the development of computer systems that are capable of learning and performing tasks that normally require human brains to do. A lot of early research in artificial intelligence involved vision, as this was one of the rare parts of the brain that was easy to understand as far as structure and operation. AI has also been used for many other specialized tasks, like speech recognition.

The idea of general intelligence in the form of an “android” or human-like robot never quite got off the ground - at least not yet. For years, this made artificial intelligence the butt of many bad jokes. It became a joke that true artificial intelligence was always “30 years away”.

However, remarkable strides have been made, even if we’ve had to redefine the expectations and roles that artificial intelligence can play in our lives, businesses, and society. Since the early years (when people imagined the HAL 9000 computer taking over the universe), AI is now viewed more realistically. It is now regarded as a tool that can help us get many specialized tasks done, often freeing up the need for human labor, in much the same way that the early machines of the industrial revolution did.

You can think of artificial intelligence as computer systems that think and learn in ways that human intelligence does. It can be specialized in tasks like

speech recognition, facial recognition, language translation, and so forth.

The development of artificial intelligence has largely gone through three major phases, although they are not mutually exclusive. The first phase involved the development of neural networks. These systems mimic the actual physical functioning of the brain. As such, computer system is divided into interconnected nodes; when the system is exposed to something, it will process the data and learn.

When a person learns something, neurons form circuits. Some neurons will strengthen their connections to other neurons, while they will weaken their connections to others. While we can describe this process, it's not understood how it works and how it is that "you" are inside your brain perceiving the information that the neurons are representing and processing. Nonetheless, neural networks mimic this structure.

An AI system that is based on a neural network will have a layer of nodes that are exposed to the data being used for learning. There are nodes that are used for output of information. In between, there are hidden layers of additional nodes. The nodes can all communicate with each other, and they have connections to each other which can be strengthened or weakened - depending on how the learning process proceeds. Keep in mind that there is no computer programmer that is directing the nodes to strengthen or weaken the connections, or to talk to these nodes over here, or those over there. This all happens independently. The computer system works and learns completely on its own. After it has been trained, it can make decisions on its own, too.

The neural network phase lasted through the 1970s. Keep in mind that neural networks are still used and it's still a hot area of research. Machine learning became the central focus of AI researchers in the 1980s, and the popularity of machine learning has continued to this day. With machine learning, which is a subset of AI, large datasets are fed to the computer system, and the machine learns from the data. This can be done using supervised learning where the machine is guided to the right answer. Alternatively, this can be done in an unsupervised fashion, and the computer is exposed to the data, and it learns on its own from the data. Machine learning has gotten extremely popular because it works, and it provides many practical applications. The third phase of AI is called deep learning. Deep learning is a form of unsupervised learning that uses neural networks.

The applications of AI are specialized, but they are no less dramatic. The

advantage of AI is that the systems can learn, and the more data they are exposed to, the better they get at what they do. AI can be used for data discovery, perform repetitive and high-volume tasks that people find tedious and boring, and to add intelligence to products that people use. This, in turn, can be used to collect even more data.

## **Data Science**

As machine learning began to enter the scene as a practical tool used by business, a new field of study emerged. This is an interdisciplinary field that combines statistics, computer science, business knowledge: data science. The job of the data scientist is to manage the use of machine learning in a business environment. The data scientist will determine what problems and dataset the machine will work with, and what questions need to be asked of the data. The data scientist will train machine learning systems and then decide when they are ready to be deployed in the real world. It is a booming field that is at the core of many trends in business today.

## **Data Mining**

Data mining, or knowledge discovery databases (KDD), is a process of finding information that is contained in large databases. Data is raw, and by itself doesn't represent information. Data mining is a process that will determine what - if any - information is contained in the datasets.

## **IoT**

IoT is shorthand for "internet of things". This is a long-promised merger of the internet with many common devices and appliances that we use. But in today's world, it is taking on new meaning, as it is going well beyond a simple internet connection. Now we are talking about ordinary devices being used to collect massive amounts of data about people in their home, or on their bodies in the form of wearable computers. There are many ways that these devices and data can be used and, of course, there is the ethical question of abuse and potential for hacking.

## **Data Analytics**

Data analytics is a process of discovering useful information in datasets. It involves many levels of data processing, such as examining the data, modeling the data, and transforming it into a more useful form. The goal of data analytics is to bring data into a form that can be used to improve



awareness among important actors and use it to support informed decisions that are actionable.

## **Cybersecurity**

Cybersecurity is the protection of computers and computer systems from malicious activity that can include hacking, remote control, intentional damage, and unauthorized access. Big data and data analytics are often used to enhance cybersecurity by using machine learning and AI systems to detect irregularities that can be associated with intrusions and attacks of computer systems.

## **Social Media**

Social media refers to a wide range of websites and apps that allow people to come together and communicate and share information. Common examples of social media include Facebook, Twitter, Snapchat, Instagram, and Pinterest. Social media represents an unprecedented opportunity to collect data on people, their associations, and their behaviors. Although search engines and shopping sites like Amazon are not social media, the combination of data from social media sites with other computer and internet-based systems is proving to be a very powerful force.

## **Internet Marketing**

Internet marketing is the process of advertising on the internet through search engine advertising, tailored ads on websites that are frequently visited, or through direct advertising to people on social media. Internet marketing uses big data, data analytics, machine learning, and artificial intelligence in order to tailor advertising to each individual. Internet marketing can use big data quickly, rapidly changing its approach in almost real-time based on recent behaviors of the consumer. If you visit a website for shoes and then start seeing ads for shoes on the Wall Street Journal, Facebook, and CNN, then you have experienced this practice.

## **Chapter 2 : An Overview of Business Intelligence**

Business intelligence is being driven by artificial intelligence, machine learning, pattern recognition, data analytics, and more. But what does business intelligence (or BI, as it is commonly known) mean to businesses themselves? For the business, the real meaning of BI is going to be the result. At the core, BI is about actionable insights from the myriads of big data that exist in the company. The data can come from many sources, and oftentimes BI is about data that is collected about the company itself! This can include information about company operations, interactions between employees, meetings, communications, phone calls, emails, and so forth. Any and all information that is collected by a business is useful for BI. Internal information can be used to massively improve company efficiency and operations, while information collected about customers can be used to improve customer service, develop new products and services, and respond in real-time to customer concerns.

### **The Key Characteristics of BI for business**

Let's begin by taking a closer look at how business intelligence can benefit a business. At the core, business intelligence is driven by data. It is also driven by the fact that modern companies need to make data-driven decisions rapidly. The goal of business intelligence is to process that data and get insights from it. This can be done with machine learning, but the goal is to provide actionable insights to the employees of the business. These insights need to be more than actionable; they need to be timely, in the sense that decisions need to be made as rapidly as possible. In today's competitive world, the competitors of any business are using these same tools to their own advantage. Those that act first – if they are acting on accurate information - are the ones that are going to come out on top. Information also *must have value*. Any old information won't do, and information that isn't valuable and accurate can lead a company astray.

Data and the information that is extracted or inferred from the data forms the core of BI. The first essential characteristic that is important for BI: the data must be genuine. Falsified or misleading data would destroy the entire enterprise. It is important for the company to ensure that it has good data sources, that the company is not being misled on purpose by other companies or bad employees within the company, and that the company's computer systems are not being attacked from the outside.

Going forward, we will assume that the data is genuine. The second key characteristic that BI must have: it must be able to provide companies with meaningful insights that have value. In order to have value, insights that are extracted from data must be able to drive important decision making. This is where the data science comes in; it's crucial to start off asking the right questions and making sure that data analytics, machine learning, and artificially intelligent systems are working on the right data that is going to be useful for actionable insights. The patterns, trends, and relationships that are discovered using machine learning need to have a significant impact on business operations. There are many ways that they can be useful:

- By cutting fuel costs
- Speeding communications between different departments and teams
- Improving collaboration
- Accelerating the delivery of products and services to customers
- Quickly identifying problems and providing solutions
- Improving customer service

Those are some of the many examples of valuable insights provided by BI. If valuable insights are not extracted from the data, then time and money are wasted, while competitors are racing ahead.

This brings us to the next key characteristic: insights gleaned from data analysis must be valuable *and accurate*. If the information that comes from the process of business intelligence is not accurate, then you can't say it is useful. Of course, accuracy in this context is not always absolute; there can be a range of accuracy. It might be accurate at a level of 50%, or it might be accurate at a level of 90%. If it has low accuracy, then it's not going to be valuable for the decision-making process in the organization. Insights that have low levels of accuracy are going to lead business operations in the wrong direction, wasting time, decreasing efficiency, and ultimately costing money. It could lead a company to develop the wrong kinds of products or to misallocate resources in the enterprise. When these problems come to light, trust in the process will be reduced, causing further problems.

The final key characteristic is that business intelligence needs to be rapid. Remember the old saying: "Time is of the essence." The faster you get information, the more valuable it becomes. In other words, the more important that something is, the timelier it needs to be. Still, it's important to

avoid sacrificing accuracy or value for the sake of getting insights out to the organization quickly. Therefore, none of the characteristics described here can be taken to exist independently. They are all interacting and crucially important, and one depends on the other and vice versa.

While we don't want to cut corners in order to achieve the most rapid delivery of actionable information possible, one cannot underestimate the cost to a business when it comes to time delays. There are many bottlenecks in an organization, and any single one of them can severely constrict workflow.

## **The Four-Step Process of Business Intelligence**

The core of business intelligence is actionable information. But how does this information get out to the organization, and how can it be made better? The way to make it better is to feed new data back into the system. This works particularly well when we are dealing with machine learning-based computer systems. Remember the checkers program (the game) that launched machine learning in the first place? The more the computer played the game, the better it got. This is true of machine learning systems that are used in business today.

The first step of the process is data collection. A key part of this process is ensuring that we are gathering the right data for the types of insights that we are interested in. Of course, beforehand we don't know what the insights are going to be, but we know what kinds of questions we would like to ask. It is going to be up to the machine learning systems to find out what the insights are, but they must be pointed in the right direction. And if the data is bad to begin with, the entire process becomes flawed.

Once the insights have been generated, the next step in the process is making data-driven decisions based on those insights. This can be anything depending on what the company is using the data for. For example, ad expenditures on social media sites could be increased. Perhaps a sales team will have its budget reduced. There are endless possibilities.

No matter what kinds of decisions are made, you must be prepared ahead of time to figure out how you are going to measure results. This is the third step of the process. The bottom line is you must know *what you are going to measure* in order to analyze results and draw conclusions from the process.

Finally, the operational results that come from the implementation of the actionable insights can be fed back into the system. The new data can be

used to enhance learning and improve efficiency in a step-by-step fashion.

## **Business Intelligence Applications**

There are many software tools that are used as part of the business intelligence process. These include some that are very familiar and mature, along with some new tools that have entered the corporate world in recent years. Let's look at some of these tools.

### *Reporting and Database Querying*

Businesses have lots of data in databases, and it's important to be able to get that data out and present it in a useful fashion. SQL, or structured query language, is a key part of that process. This allows you to get data out of large databases, search it, and organize it in useful ways. You also need a way to present data in a readable format to human beings. Reporting software that works with company databases serves this purpose.

### *Spreadsheets*

Data can be drawn together or filtered and presented in spreadsheets. Ever since their introduction in the early to mid-1980s, electronic spreadsheets have played a fundamental role in gathering data and turning it into a useful form. In case you mistakenly think that modern systems are making spreadsheets less relevant, *this is not the case!*

### *Data Dashboards*

Dashboards are viewed on a web page. A dashboard is a computerized interface that provides the users with real-time reports that are updated in real-time using a database. The application of dashboards is widespread and can be used in any business application in which seeing real-time updates of data is useful, such as viewing inventory, staffing, or manufacturing data.

### *OLAP*

OLAP stands for "online analytical processing". OLAP is a set of tools that helps users analyze multidimensional data. OLAP is used for business reporting, budgeting, and forecasting.

### *Data Mining*

Data mining is a process by which large datasets are analyzed in order to detect patterns that exist in the data. It uses methods from machine learning, statistics, and from databases when large amounts of company data are stored in relational databases. The main goal of data mining is to extract data and

transform it so that it can be presented in a useful and readable form. Data mining also seeks to reveal patterns that are in the data and extract useful knowledge.

### *Data Warehouse*

A data warehouse is a computer system that serves to perform data analysis and reporting that is useful to the enterprise. A data warehouse also centralizes data collected from many different sources.

### *Data Cleansing*

Data cleansing is a supportive but very important task in terms of maintaining the data that is used as a part of the business intelligence process. It involves the detection, repair, or deletion of corrupt database records. Data cleansing is necessary in order to continually ensure that the data used by a company is accurate.

### *Business Activity Monitoring*

Business activity monitoring is a software system that uses data dashboards to present important information on the activities, operations, and processes taking place in an organization. The goal of business activity monitoring is to help an organization make more informed decisions that are data driven. It can help companies quickly identify problems within the organization and reallocate resources as necessary.

## **Real-Time BI**

As computer systems continue to develop and incorporate the use of mobile technologies, the potential for using real-time data in business operations is becoming more important. Real-time data can be presented to key personnel involving business operations, inventories, advertising and marketing spending, social media contacts, email, and messaging. Data can be collected in real-time to quickly launch marketing campaigns or adjust existing advertising efforts, or to contact customers directly via email, text message, or on Facebook with tailor- crafted offers.

## **The Core Benefits of Business Intelligence**

Business intelligence is a set of tools that can help managers and business leaders act on what the data is telling them. This can help the business respond in a timely fashion, rapidly responding to changing market conditions and fixing internal problems to improve efficiency and gain a competitive edge rapidly. Data-driven decisions allow business leaders to

take a proactive stance allowing the company to become nimbler and more responsive. BI will allow the company to gain unexpected insights into their data and take the company in new and powerful directions that they would not have done otherwise. The business environment is fast-paced and dynamic, and business intelligence is going to help companies navigate this environment as it becomes even more complex, and the pace quickens. With solid data, business intelligence will also help companies make fact-based decisions, whereas in the past, they were operating in the dark. Those companies that don't utilize the tools of business intelligence are going to find themselves falling behind.

## **Past or Future Data?**

One of the issues that are important to understand about business intelligence is that it relies on using *past* data. As such, it is interested in analyzing historical data and presenting it in an easily digestible form to members of the company. Then company members can make actionable decisions based on the data and its analysis. This can be in the form of readable reports and visualizations. Visualizations can include graphics, charts, and so forth that are readily understood by human readers.

Many other types of analysis associated with big data - in particular, machine learning and data analytics - are used for the purposes of *predictive modeling*. That is, think of business intelligence as backward-looking, gleaning insights that can be used by humans to make business decisions. Other types of data analysis are going to be used to take future action or predict what different courses of action will produce. We will discuss these issues as we go forward, but one might consider that as more tools become available that utilize the ability to act in future situations and predict the what the results are going to be based on new input data, it is highly probable that business intelligence is going to evolve to incorporate these tools. You are probably going to witness business intelligence incorporating machine learning in order to develop self-service BI.

## Chapter 3: BI and AI

Data and insights *from data* are at the core of business intelligence. This is where artificial intelligence and machine learning come into play. AI and machine learning help extract meaning from the data that is so crucial for business intelligence. In fact, without machine learning and artificial intelligence, business intelligence would be years behind where it is today. In this chapter, we are going to discuss the role that artificial intelligence can play in the growth and enhancement of business intelligence. Some readers may find the terms “machine learning” and “artificial intelligence” confusing. Here’s a distinction: machine learning is a subset of AI and has specific uses that we will discuss in this chapter.

### The core uses of AI

Artificial intelligence is going to be used in a wide variety of ways. You already see early applications of artificial intelligence in the form of self-driving vehicles, AI systems that can analyze MRI scans for signs of tumors. Language translation services are becoming nearly as accurate as a human translator. AI is used in speech recognition systems, facial recognition systems, and more. AI is at the beginning stages of development, and so the applications are only going to increase. In this section, we are going to look at the core ways that artificial intelligence is used and how it can be applied in business.

One of the ways that AI can be used is to add intelligence to existing systems. That is, AI will be added to existing products in order to transform them from dumb inanimate objects into more useful intelligent systems that can better serve the needs of their users. One area where this is already being done is adding intelligence to automobiles, such as the case of regular vehicles that are not self-driving cars. Adding limited intelligence capabilities can help drivers operate the car more safely and improve the odds of avoiding accidents.

Soon, AI can be added to a wide array of devices to connect them to the internet, collect data, and then provide upgraded capabilities as a result of making currently mundane devices useful and intelligent. For example, a toilet can use AI in order to detect early signs of disease and communicate this information to the doctor. AI can also be used to add conversational capabilities to products, better serving the needs of customers. For example, an intelligent refrigerator could let the owner know that supplies of various



foods are running low, and the device could even order the food items via an internet connection. Smart machines can also be used for security or monitoring public spaces for safety.

Data processing is another area where AI can be used, and it is already being used extensively via machine learning. This is typically used to spot patterns in large datasets; that information can then be used in many different contexts by businesses. Insights from this kind of analysis have been used to improve the efficiency of logistics, cut fuel costs, improve the allocation of staff and other resources, fraud detection, and approve customers for loans and credit, among other applications. The use of machine learning is sure to grow in the coming years and become more widespread in the business community. The important thing about machine learning is that it can extract insights from large datasets that could not have been extracted in any other way.

Machine learning can also be used on internal data collected by the company. Often this kind of data is unstructured and in the form of images, recorded phone calls, and emails. Artificially intelligent systems can analyze this data for unexpected patterns that often produce actionable insights that help the company organize itself better and more efficiently.

Neural networks using deep learning are being used to detect unexpected or anomalous patterns, improving fraud detection, and cybersecurity. The increased levels of data combined with better neural networks are making these processes more useful and efficient. Like the original checkers-playing software, these systems are constantly learning and improving their skills as they are presented with more data.

## **AI Can Make Decisions on Its Own**

Business intelligence is about gathering, processing data, and getting actionable insights from the data so that it can be presented to humans, allowing them to make decisions based on the data. Artificial intelligence adds an interesting wrinkle to this model, because AI systems - once properly trained - can make decisions independently from human supervision. AI systems are already being deployed in many contexts to perform work and make decisions that humans used to make. For example, they are used to generate credit scores and analyze loan applications, or they are used as chatbots to provide the first level of customer support. These systems work autonomously without human intervention.

The first question that comes up in this context is: How is AI going to impact

BI? Let's think about it. One benefit of AI applications is that this frees up humans for more interesting work. AI is taking over the tedious types of tasks – many of which humans couldn't perform anyway. Think of big data in this context. A human mind is not even capable of examining big data and extracting insights from it. So, in a sense, a new job has been created for AI to perform that didn't even exist without AI. In a lot of cases, AI can help an organization become more efficient by removing the need for people to be involved in many tasks that rely on tedious data analysis, moving humans into roles that are more valuable because they require insight and they are far less tedious. Think about how AI is automatically enhancing the security of the organization by being able to search out anomalous behaviors on computer networks in order to provide better cybersecurity.

According to the Harvard Business Review, the use of AI in business falls into three general categories. Let's look closely at each of these.

## **Process Automation**

Process automation is becoming a very common application in businesses increasingly driven by AI. Process automation involves using intelligent computer systems to perform administrative tasks in a business. These can include (as we noted above) handling frontline communications with customers, updating databases, sending out bills, and recording documentation.

## **Cognitive Insights**

This is a more advanced use of AI in business. These are machine learning systems that become better at their jobs as time passes because they can learn from data that they encounter continually. These types of systems are used for pattern recognition, development of tailored advertising, cybersecurity, and to predict customer behavior.

## **Cognitive Engagement**

These types of systems are designed to interact directly with the public, and with employees. They can include intelligent phone systems that offer customer support or take bill payments, chatbots, customer service, or systems that give medical advice. AI systems can also be used to develop personal health coaches, reminding people to take their medications, examine dietary choices, and making sure people have kept up with their exercise plans.

## How AI will Take BI to a new level

The purpose of business intelligence is to extract meaning and insight from a company's data. Artificial intelligence is going to enhance this process massively. AI is not going to replace all people, but it will help to process the large amounts of data that companies are collecting and turn it into useful information. AI will also play a role in analyzing that information and synthesizing it to help humans make better decisions based on the data. AI will also be able to play a role in assessing the integrity and value of the data. The higher speeds at which AI and machine learning systems can operate allows them to analyze company data much faster and better than any human. This can mean that AI systems can synthesize data and do preliminary processing, saving people from having to do a lot of tedious work.

Another area where AI is going to be used more and more is in the area of improving business intelligence with the use of *predicting modeling*. This can be incorporated into business intelligence to predict the outcomes of different courses of action before implementing them. Then, using the results from such modeling, the best solution that is identified can be the one that is put into practice.

## Chapter 4: BI and Big Data

In the modern era, big data is playing an ever-increasing role. Corporations have collected massive datasets about their customers and about their own internal operations. With machine learning, businesses can extract valuable information about the data that would otherwise be impossible. As explained earlier: human beings lack the capacity to analyze large datasets and find relationships in them. But for artificially intelligent machines, this is an easy task. Not all patterns, trends, and relationships found in the data are going to be useful for the company - or even cost-effective. The role of business intelligence when it comes to big data is to determine whether it's actionable, accurate, and valuable information.

### Big Data Classifications

There are three general classifications of big data: unstructured, structured, and semi-structured. All three types of data can be mined for information that might be of use.

### Structured Data

Structured data is *organized data*. It is stored in relational databases and other organized data structures, and it's based on a data model that can represent something like a product in inventory or a customer. The data model will flesh out the properties that the company wants to track. An example of structured data would be an employee listing in a database that had the employee's name, age, salary, home address, and so on. The elements, or records, in a relational database have addresses that can be used to look up individual records easily. Structured data can typically be queried using structured query language or SQL so that subsets of the data can be extracted from the database. The data can be sorted, grouped, or filtered as desired. Structured data has been with businesses and used in business intelligence for several decades now. SQL was developed by IBM in the early 1970s, and over the following decades the use of relational databases became more common, growing in sophistication and usability. As soon as data was organized into databases like Microsoft Access, programs were built to utilize it. Information from databases is easy to analyze and present in the form of reports or on data dashboards.

Each element in a record in a structured dataset is called a field. Fields can be one of several basic data types such as alphanumeric or text, numeric, or

currency. An employee's name or their social security number would be a field. Data can be queried from multiple databases and combined into single records using a runtime command, providing new information that could be useful to the business. For example, a business could have a database of customers and a separate database of past purchases by a customer. These could be combined to yield useful information about each customer or about products. For example, if the company had each customer's address, it could pull up reports that showed how each product sold in different locations.

Spreadsheets represent another form of structured data that has also been a critical component of business intelligence. A spreadsheet has the advantage of providing a unified platform that can contain the data, together with an analysis of the data and graphical presentation. Analysis can be changed in real-time to get new insights from the data. Spreadsheets arose from older methods of working with data on paper, but soon revealed the exponential increase in power that they offered when looking at data and gathering insights from it. As soon as computerized spreadsheets became available, companies flocked to them in large numbers. The first commercially available spreadsheet programs were made in the early 1980s; by the end of the decade, nearly every business was using them.

Structured data has many characteristics that have made it easy for humans to work with, including easy data entry and the ability of computers to quickly pull up data and organize it into useful forms such as graphics, pie charts, histograms, or reports.

Even though structured data is highly organized and easily interpreted by humans, there can be insights into the data that humans are not able to extract. The reason is that the datasets have grown incomprehensibly large, and there are large numbers of ways that related databases can be combined. Once combined, each new combination of data yields an exponentially number of useful insights for the business. This is where machine learning has come into play. Before the past decade, the processing of structured data was done with "dumb" computers that processed data without thinking about it. These were conventional computer systems that only implemented the instructions that were given to them by computer programmers, and they followed a pre-determined set of rules. Those types of computer systems are good for mundane tasks, but they are not able to think, interpret, or act autonomously in the way that artificially intelligent systems can.

# Unstructured Data

Big data can also be unstructured. This is information that is not stored in a relational database or some other data structure-based form such as a spreadsheet. You can consider this type of data as the opposite of structured data. It is not easily addressable by SQL or pulled together and analyzed by humans.

Today, most data collected by businesses is unstructured data. It is estimated that about 90% of the data collected is unstructured, and this percentage is likely to increase in the coming years. This is going to make machine learning more important for the business since it plays a central role in analyzing unstructured data.

The business has many forms of unstructured data. This can include multimedia content such as images, videos, and even slide show presentations. Although you may think of text as structured data many text-based documents are unstructured data. This can include social media messages, text messages sent over mobile networks, emails, memos, and word processing documents. Images, like a word processing document, are highly organized in and of themselves. However, a collection of millions of images or word processing documents that a business may have stored is completely unstructured. So unstructured data has internal structure, but the collection of data is unstructured. The collection of data would be difficult if not impossible to organize into a database or spreadsheet, and even if it could be organized in such fashion, it would be extremely difficult and time consuming to extract information from it. The internal structure in unstructured data such as emails or images can contain extremely useful information.

You can imagine the problems that would arise while using humans to try to organize and analyze unstructured data. A program like Microsoft Excel would not be capable of analyzing images and organizing them as a result, much less finding actionable information from the images. A human being might be able to do it, for example, by examining each image submitted by a user on a social media site, classifying the image in some way. However, to go through large datasets, a human being would take thousands of years or more just to get through it. Mistakes would be made, and insights would be missed. Human beings would also suffer from fatigue, slowing any analysis of such data and causing even more mistakes.

Gaining insights from the data and relating different data from the same

person (for example, images uploaded, and text messages sent) makes this task utterly impossible for human analysis and even traditional analysis used with structured data.

Until the development of machine learning, effectively working with unstructured data was impossible. Machine learning allows data to be mined, processed, and analyzed to find hidden patterns that can be used to create actionable insights for business.

This is an important development for businesses because large data stores of unstructured data contain a lot of information which can lead to better and more informed business decisions. Since the internet became widespread in the late 1990s and computer storage capacity increased, the collection of unstructured data by companies of all sizes has grown exponentially, and it continues to expand with each passing year. In fact, the pace of collection of unstructured data is quickening, making the ability to process it and analyze it very important for business intelligence.

Just searching through the data is a challenge in and of itself. Software tools have been developed that help companies mine and organize the data. It can then be presented to machine learning systems that can glean actionable insights from the data.

Unstructured data can come from multiple sources. When people think of unstructured data, an example that comes to mind is social media. Each individual or business using social media is creating a large set of unstructured data that includes images, videos, tweets, postings, comments, or messages sent to other users.

However, there is also a large amount of unstructured data that is being collected internally by businesses on their employees. This includes emails, memos, word processing documents, video footage of facilities, and even logs of computer usage. This information - if it could be analyzed - would provide an endless array of actionable insights for business intelligence that could be used to improve the efficiency of the company.

These benefits can work across a large spectrum of applications. For example, one use of logging computer time is to identify employees who are unproductive or who waste time at work by logging into Facebook or spending too much time reading personal emails or checking favorite websites. This allows the company to weed out unproductive employees - or at least give them a chance to adjust their behavior.

Tracking of employees is now possible, so companies can know where

employees are at all times and how they are spending their time inside company facilities. However, it's not enough to just have the tracking data. The quantity of the data can explode very quickly. To become useful, the data must be gathered, analyzed, and then interpreted so that it can be used with business intelligence to improve operations.

The list of possible unstructured data sources is endless, as are the possibilities of gleaning actionable information. For example, emails between employees can contain a treasure trove of information. By analyzing emails, a company could find better ways to organize staff and encourage collaboration to improve efficiency. It can also be used to spot waste, fraud, and abuse that could be occurring inside the organization.

The key to using unstructured data and getting actionable insights from it is machine learning.

## **Semi-Structured Data**

The third class of data exists that is in between structured and unstructured data. This is data that is a mixture of structured and unstructured data, or at least it contains data that could be organized as structured data.

## **The Five V's of Big Data**

When people talk about "Big Data," they often talk about the five V's. In fact, in recent years some have expanded this to the six V's. These are characteristics that describe big data. These include:

- **Volume:** Big data is a lot of data. It can be collected from multiple sources.
- **Velocity:** Big data is being generated at an increasingly rapid pace.
- **Variability:** Big data comes in a wide variety of forms, and the ways that big data can be used are infinite.
- **Variety:** This refers to the types of big data described in the last section – structured, unstructured, and semi-structured data.
- **Veracity:** How reliable is the data? This is a measure of how much the data can be trusted.
- **Value:** The final "V" on our list is the key one for business intelligence. This is the value of big data, and whether it is actionable for the business.

Keep in mind that not all data has value. In fact, whether data can provide



actionable insights doesn't just depend on how revolutionary the discoveries made from the data are. There are going to be many practical considerations. For example, an amazing insight discovered by AI about data a company has would not have value if implementing any actions based on it are not cost-effective. Another possibility is that it may be impractical for the company to carry out.

Nonetheless, it's clear that companies that are analyzing their big data are going to be the companies that gain a competitive advantage. Although some insights won't be actionable from a practical standpoint, many will be. This will lead to better relationships with customers. It will help the company allocate resources better. Companies using big data will be able to develop better products and services, and they will be able to respond to maintenance and support issues with targeted efficiency. In addition, they will be able to restructure their organizations in real-time in order to massively improve efficiency, collaboration, and throughput. In short, they will gain massive competitive advantages – and we see that big data plays a central role in business intelligence.

## **How do Big Data and Business Intelligence Compare?**

This is a question that often comes to mind among people learning about these topics. However, the question comes from confusion about what big data and business intelligence are. Business intelligence is the complete set of data, tools, insights, and actions that arise from the data that a business has access to. Business intelligence leads to data-driven decision making.

Big data is nothing more than data, and it's not useful unless it can be searched, organized, and analyzed. So, you cannot compare big data to business intelligence.

Big data is a component of modern business intelligence. If you recall from our earlier discussions of business intelligence, it begins with the data; this is where you will find the connection between big data and business intelligence. It is a part of the foundation upon which business intelligence is built. In fact, today, you can say that for many companies, big data *is* the foundational component of business intelligence.

## **Big Data Resources**

Big data can be an expansive resource for companies. This is because many

sources of big data exist, and they are accessible to businesses as big data becomes the new currency. Therefore, a company is not limited to the data that it collects on its own. Big data is a large component of buying and selling between businesses these days and has been for decades. In past years, businesses could buy customer lists from other businesses. Today, you can buy big data and even have it analyzed for you, as we will show with our case study for small business below. The key point here is that businesses should not be limited in their thinking process about data. They should recognize that data sources exist for them inside and outside the company.

## **Developing a big data strategy**

A big data strategy is going to involve the collection, storage, retrieval, analysis, and actionable insights. The goal of a big data strategy is to build a big data foundation that can power business intelligence.

The collection is not an issue for most businesses, but many businesses err by pre-judging the value of data that they could collect, thus not collecting as much data as they should. By now, you should recognize that data may have insights that are not readily apparent. This is the benefit of machine learning and the role that it plays in data analysis: finding hidden patterns that humans didn't know existed and - frankly - couldn't have known existed. Therefore, you need to collect data with an open mind. Any bit of information that a business can collect is potentially useful. At first glance, it might appear of no consequence, but it might be used later to draw unexpected connections that can lead to data-driven decisions later.

The second issue that needing addressed as part of a big data strategy is the *storage and retrieval of the data*. A large amount of infrastructure is needed to handle the volume and velocity of big data. This includes large storage capacity as well as powerful computers. It is difficult for organizations - especially smaller to mid-sized businesses - to build the kind of data-driven infrastructure that will enable them to handle big data effectively. Doing so could literally require hundreds of computer servers with huge banks of data storage and highly efficient data retrieval tools. In addition, a high level of cybersecurity is necessary to protect the integrity of the data.

For many businesses, these requirements mean that they are going outside the organization to have these needs met. Cloud computing has made this possible, and many of the world's largest companies are offering accessible cloud computing resources that are cost-effective, fast, and easily used to

store and retrieve data on a 24/7 basis. Providers of such services include Amazon, Microsoft, and Google. These companies offer several levels of cloud services that can be tailored to fit the needs of any business size, from a one-person operation running as a home business up to a large corporation - or even government entity. For example, Amazon's Simple Storage Service, known as Amazon S3, can be used by businesses of any size for low-cost cloud computing services.

Companies that can deploy their own big data systems can utilize open source tools to get the job done. This includes using Apache tools such as Hadoop and Spark, along with MapReduce. Careful considerations must be made. Handling the job internally will keep the data under the direct control of the company. However, this will entail having to build out or rent massive computer storage capacity that must be maintained on a 24/7 basis. That will also require a robust set of information security measures to ensure the integrity of the data and protect it from hacking, corruption, and other problems. This will also require hiring a large amount of staffing.

Even though many larger businesses might have the capacity to implement their own completely internal big data strategy, it's not clear that doing so is always the best way forward. This decision will have to be made by each organization. However, the fact that established companies like Microsoft and IBM are available for providing infrastructure and services that are already thoroughly tested and robust makes outsourcing these tasks a viable and more cost-effective option.

By utilizing third-party services, you can massively reduce the cost of implementing a big data strategy. Big data capabilities will be provided for you by established companies that have already developed massive computer infrastructure for storing, maintaining, and securing data. Third-party sources like IBM, Google, and Microsoft may also be able to assist with analytics.

Once the collection, storage, and retrieval problems are solved, systems must be put in place to make sense of the data. By itself, the data is useless. To be used for the kind of data-driven decision making that business intelligence entails, the patterns, relationships, and trends in the data must be discovered and put in a presentable form for human analysis. These steps will involve the use of data science and machine learning. This is another step where a company will have to decide whether it wants to do this internally or outsource the task.

There are several arguments that can be made for outsourcing the task for small and mid-sized businesses. Many companies like IBM have a long history of developing artificial intelligence and machine learning tools, and they are making their capabilities available as business services they are providing for clients at an effective cost. IBM has been a leader in this area since artificial intelligence first became a field, so you know you are getting high levels of reliability. This is not a recommendation for or against IBM, but rather to let you know that these types of services are available. They are also available from other large companies like Microsoft, Facebook, and Google, as well as from many newer and startup companies.

Many large corporations are doing this internally. Southwest Airlines is one example.

Either way, a part of your big data strategy is going to include developing a data science team. If you outsource your big data analysis, you may still need an internal data scientist to collaborate with third-party teams to get the most out of the data. Data scientists can be hired internally, or you can work with contract providers.

Machine learning tools for doing data analysis is also available for internal use, and you can hire data scientists to develop their own tools. Depending on your specific needs, there may be off-the-shelf software tools that you can use, and they are already proven to be reliable and effective. For this reason, you may not necessarily want to hire a team of data scientists to develop your own internal tools. Of course, if the expense can be managed and there are special needs in a company that cannot be addressed by off-the-shelf components, this can be an option to pursue as a part of your big data strategy. Another alternative is to utilize existing services from companies like IBM, Google, and Microsoft to turn your big data into insights that can be incorporated into business intelligence.

The final part of a big data strategy is engaging in data-driven decision making based on the actionable insights that have been derived from this process. This is going to become a central part of the business process, operation, and organization. This is because it is a continuous process. Machine learning is not a one-time event; it is going on constantly as large volumes of data continue to be collected. It operates in this way to be able to provide new insights often on a real-time basis after the infrastructure has been established.

In each case, the actionable insights provided by this process are going to

provide a basis for business intelligence as well as provide inputs for business intelligence. Human insight and ingenuity will be brought to bear in many cases. Not every insight is going to have an equal value to the enterprise. Therefore, all the existing components of business intelligence will have to be applied in order to determine which actions can and should be taken by the organization. Issues such as cost-effectiveness are going to be central, as well as the ability of the company to carry out each action from a practical standpoint. The allocation of limited resources to different possible actions is another factor. The leaders of the business at all levels of management will have to weigh each possible action against others. This is nothing new, and businesses may have to forgo some actionable insights even if they are valuable in favor of others that are going to be impactful and be more cost-effective. The situation will have to be evaluated and re-evaluated on a consistent basis as more data continues to come in.

Not all possible actions that come out of the application of machine learning and AI to big data are going to require the input of humans. In many cases, these systems will be operating autonomously and without human input, other than a periodic review. Examples of these types of applications include cybersecurity and fraud detection in financial matters.

Many people new to the concept of machine learning are hesitant – even put off - by the autonomous capabilities that it offers. However, this can be a huge competitive advantage to the organization. Rather than making humans useless, it frees them up so they can be used for higher-level and more important tasks. It also helps the company work far more efficiently, both internally and externally. An example of this is frontline customer support operations. By using an AI system to handle this, employees can be dedicated to handle customer problems that are of a more serious and impactful nature. This will improve the ability of the company to respond to customer complaints and problems by providing better service. However, one side-effect of this: companies and society at large will have to devote more resources to employee training so that they are able to perform at a higher level that requires a larger and more sophisticated job skillset.

## **Problems with Big Data**

The value of actionable insights arising from big data is only as good as the original datasets themselves. As they say in computer science: “Garbage in, garbage out” (GIGO). Errors in big data can occur at the collection point of

the data, but data can also be misinterpreted. But, for businesses to have the most efficient and beneficial big data strategy that they can possibly have, they need to begin at the point of data collection.

In many cases, there is not enough data collected, as paradoxical as that sounds. While companies may have large amounts of data, they may also be trying to get insights from *data points* that are not known in enough detail in order to bring actionable insights.

Analysis of big data also suffers from a problem referred to as the *bias-variance problem*. This is closely related to machine learning, and we will discuss this problem in a later chapter. But be aware that the bias-variance problem - if not handled correctly by data science professionals - can lead to erroneous conclusions.

Another problem that can plague big data is that large errors can occur when fictitious statistical relationships are derived from the data. Since people have great faith in the infallibility of computers, they often ignore the fact that bad data could have been inputted to the computer system, or the data may have been presented or analyzed in the wrong way. This can lead to data-driven decisions that are based on completely or partially incorrect insights. This mistake can be costly for businesses and decrease competitiveness.

Often, the data is good. But one problem with large amounts of data is that it will always include noise and outliers; both can lead to erroneous insights. When data is presented to people, they can sometimes be led astray by outliers that are not truly representative of the data. This is a problem that happens all the time.

Errors are common with the applications of big data, and machine learning systems can often misinterpret the actions of people. Consider the behavior of people online. Specially tailored advertising is all the rage among businesses these days. However, something that often occurs is that someone may look at a product or article for reasons not related at all to their own consumer behavior. For example, a friend may ask you to look something up. But the AI systems that are collecting data about you may not know about this, and they will make the mistake of believing that you are looking it up because you are interested in it. Or maybe you read about something in an article, and then look up a related product on Amazon based on nothing but curiosity from the article. You will soon be bombarded by advertisements that you have no intention of clicking on, much less acting on. In this sense,

machine learning and AI systems are actually “dumb.” Businesses need to familiarize themselves with weaknesses as well as the benefits of big data. Unfortunately, people are so enamored with the concepts of big data, machine learning, and artificial intelligence, that they are only willing to see the upsides while ignoring potential problems. Keep in mind that the cost of the problems can be huge if they are ignored.

Focusing only on the upsides can lead to a business having false confidence in the insights derived from big data.

One way to address many of the errors that arise from big data is to limit the sizes of datasets that are exposed to computer systems. The larger the dataset, the more likely it is that it will contain frivolous and erroneous data. Noise increases with the size of the dataset, and the larger the dataset, the more likely it is to contain outliers that are not useful for actionable information.

Another issue related to problems with big data is integrity and security. The way that big data is derived is important. A business needs to be certain that the data they are using is derived from reliable and accurate sources. Although big data seems to create the appearance of enhanced reliability – and generally it does – technology can cut multiple ways. One development that can arise from the improvements in technology is that malicious actors also have many toolsets they can use. The obvious example is hacking to expose data. However, malicious actors could also create large amounts of false data that could be used to mislead competitors. That data may be “available” to companies from external sources, or a company could have its own data compromised and corrupted, and the company may not even be aware of it. This can be costly down the road if false, misleading, or corrupted data is used to drive decision making. Machine learning and AI systems may not be capable of detecting false or misleading data. Their insights, once again, are only as good as the data that starts the process.

One result of this is that cybersecurity is going to be playing an ever more important role in business intelligence through its use in ensuring the integrity of data that companies have access to.

Another error that can occur is from the sources of data that could be fake. This is even a problem on social media sites. For example, many large Twitter accounts have fake followers. These followers can be robots that even post messages on these accounts. Many Facebook profiles are fake, and even use real people - or images from real people - to create their entirely

fake profiles that can be used for malicious purposes. These issues can create problems when it comes to collecting and interpreting a large amount of unstructured data that is available through social media. The ability of social media companies to go through this data and identify false and misleading data is going to become more important as the years go by.

## **Small Business Case Study: Advertising Using Lookalike Audiences**

Many small business owners feel overwhelmed by the concept of big data and feel like using big data, and the power of machine intelligence is beyond their reach. However, this is false, and in many cases, it's easy to leverage the data and machine learning capabilities of large corporations. One prominent example of this is lookalike audiences used for Facebook advertising.

Facebook advertising is a huge resource and game-changer for small businesses. By making its data available to any advertiser, Facebook enables small businesses to reach and target prospects in ways that were never possible before.

However, it goes beyond that. Small companies can exploit their own data, as well as data available from third-party sources by using the machine learning capabilities that Facebook has and makes available to advertisers.

Let's say that you are selling a product that would be used for horse training. You can collect a large amount of data from this. For example, you can run advertisements on Facebook, and while – at first – they may not be that effective, over time you will collect a database of users that have responded to the advertisement. You will not know much about the users, but you will have enough information that can be exploited by Facebook in order to get amazing insights.

In addition, you will probably be able to obtain data on other customers from other companies that have purchased horse training products. These can be used separately or together with the data that you have collected. Keep in mind that we are not talking about using data in the old way. In the old days, if you were able to obtain a database of people who had purchased horse training products, you could try marketing your product directly to these people. However, in the age of machine intelligence, we are going to take a different and more effective approach.

This data can be uploaded to Facebook. Then, Facebook will use its machine



learning tools to analyze the data. It will discover hidden patterns, relationships, and trends that are in the data that might be completely invisible to human observers. Once it finds out these patterns, it will then use them to build a database of people for you from their general audience. These people will be connected by the hidden patterns that are present in the data whether they have purchased a horse training product or have not. This happens because they share the hidden characteristics that have been discovered through the application of machine learning; they are extremely likely to purchase horse training products.

The database that they create for you, above, is called a lookalike audience. Once it has been constructed, the small business can then run advertisement campaigns to this lookalike audience. These campaigns are far more effective than traditional advertising.

This example shows how a small business can leverage the power of third-party services to use big data together with machine learning, without having that capability themselves.

## **Chapter 5: BI And Machine Learning**

Without machine learning, the big data that is now available to businesses, and that can be used to drive business intelligence, would be virtually useless. It is machine learning that converts reams of data into useable insights that can support data-driven decision making by human elements of the enterprise. In this chapter, we are going to take a closer look at machine learning so that you will get a better understanding of what it is, how data is used in conjunction with machine learning and the applications that machine learning has in the real world. When you understand these issues, you will be able to grasp the importance of machine learning as a focal point in business intelligence going into the future. The fact is that businesses that effectively leverage machine learning are going to have large competitive advantages in tomorrow's marketplace.

### **What is Machine Learning?**

Machine learning involves computer systems and intelligent algorithms that can learn from the data they are exposed to. When a computer system is using machine learning, it does not need to be explicitly programmed, and it is generally not acting according to pre-defined rules telling it what to do. Moreover, machine learning systems continually learn from new data that they encounter. In this way, they act as human intelligence acts, refining themselves when learning more from every new exposure to data. This allows machine learning systems to improve their performance with time, even after being deployed in real-world applications. This type of development hasn't been seen in actively deployed computer systems before. Two examples of where this learning ability in real-time is very useful include cybersecurity and fraud detection. The more data that the systems are exposed to, the better they get at finding a fraudulent activity or increasing the security of information systems that are under their watch. It is interesting and important to note that this all takes place autonomously and without human input once the systems are deployed and in use.

### **Classifications of Machine Learning**

There are several major classifications of machine learning, but you should be aware of them without boxing in your thinking. We begin by considering regression problems. These types of problems can be handled using algorithms based on an appropriate type of regression. The parameters used

in the algorithms will be discovered through the exposure of the system to data. Examples of general applications when it comes to regression problems *include predictive analysis and forecasting*. These abilities are completely general and therefore can be applied to a wide array of specific applications. For example, forecasting can be used to predict the failure of a key component in a manufacturing facility. It can also be used to predict future market decisions or to predict the behavior of customers that deal with the company. Machine intelligence, when applied to forecast, can also be used to develop more accurate weather forecasting and to make predictions for stock market behavior and to invest.

The second type of application for machine learning is called *classification*. Once again, this has a broad scope. There are many classification problems that are suitable for use with machine learning. One of the earliest applications of machine learning to classification problems was the use of machine learning in spam detection. While the system is still not perfect, it has greatly improved the robustness of email systems, allowing users to clear their email inboxes of unwanted spam emails. Spam detection is considered one of the classic examples of the classification problem.

There are many other types of classification problems that can use machine learning. These include image analysis. For example, you could use image analysis to classify images as to whether the image contains a person, a dog, or only material objects. Medical diagnostics can sometimes be framed in terms of a classification problem. By presenting X-ray or MRI images to computer systems, images can be classified based on whether they have a tumor in the image.

Another application of classification is fraud detection. By studying outlier behavior, a machine learning system can classify network access attempts as valid or invalid or look at debit card usage and automatically shut off a debit card that shows evidence of fraud. Although errors sometimes occur, the benefits far outweigh the inconvenience of occasional misidentification. And when that happens, remember that machine learning algorithms are not static. Like people, they are constantly learning, and so they learn from their mistakes, and their ability to correctly classify improves with time.

The next type of machine learning you need to be aware of is called *clustering*. Often data points can be clustered together based on certain characteristics. If you have a large and diverse population in a city, although people may be living in many different areas, you can analyze the data and

regroup it, say, by education level. The data can then be used to identify shared characteristics that people with certain education levels may have, and then develop new algorithms based on these insights. Clustering can be applied to different applications with a wide range of complexity.

A similar example of clustering from a completely different application could be image analysis. Pixels are naturally distributed by location within the image, but using clustering, we can group pixels together by color instead. This ability can lead to speedier information processing, since actions on specific clusters of data can be completed faster once they have been identified. Thresholds are often used to determine how close a given object in the dataset is to a cluster, and a distance function can be used to quantify that closeness. Clustering has a wide range of applications, including product recommendations, as well as database and customer segmentation.

Next, we come to *reinforcement learning*. This type of learning involves the distribution of rewards, and this can be done with computer systems or humans using computer systems. This can be used in real-time decision making and marketing. For example, reinforcement learning can be used to send out coupons to customers at appropriate intervals. One big application of reinforcement learning is in the design of video games. By giving rewards at appropriate intervals, the video game creator can keep players in the game longer, enticing them to come back to play the game repeatedly. AI used in video game development is proving to be quite effective. Reinforcement learning can also be used on AI systems themselves, aiding them in skills acquisition.

Next, we come to *dimensionality reduction*. Often, there are too many features of any object when it comes to machine learning. The number of features that characterize an object is often referred to as *the dimension*. While large numbers of dimensions contain a large amount of information, higher numbers of dimensions also make it more difficult to arrive at that information. By using dimensionality reduction, problems can be made more practical, allowing machine learning systems to be more effective at accurately identifying hidden patterns and relationships that occur in the data.

## **Training Machine Learning Systems**

As the name implies, a machine learning system must learn from data. There are three ways that an AI system can learn from data. These are called supervised learning, unsupervised learning, and semi-supervised learning.

We now consider each in turn.

## Supervised Learning

Supervised learning is used when there is already something known about past data that has been collected. In this case, there are known outputs for given inputs in the data. This type of data is well-suited for use in classification problems or with regression. For example, spam email can be classified as spam or not-spam based on characteristics of the email, and there is a large amount of past data that can be used to verify these facts.

Supervised learning, therefore, is used to map inputs to outputs. We say that the data used to train the artificially intelligent system is *labeled*. Labeled data allows us to present the system with input vectors, which are collections of features that characterize the object of study. Vectors can contain any number of dimensions, but one of the challenges for researchers and engineers is getting the number of dimensions right for each problem they encounter.

By exposing the system to input vectors with known outputs, the system can derive the relationships that exist between the input vectors and the known outputs. The goal of this kind of training is to help the machine learn enough so that it can develop the correct algorithm that describes the relationship between the input vectors and the outputs. This relationship may be completely unknown to humans, and in fact, it is probably unknown, because if people know the relationship between inputs and outputs, they can explicitly program a computer to implement that knowledge in an algorithm that can be applied to new data. So, in this situation, you are not going to know what the relationship is, but if the training is done correctly, the machine learning system will discover it.

Once the system has been trained, it will be able to correctly derive outputs for new inputs that it has never seen before. When engineers have the assurance that this process is working correctly, the system can be deployed in the real world, where it will presumably correctly identify outputs from new inputs that are coming in with new and previously unseen data. Remember that these systems continue to learn once they are deployed, so the ability of the system to do its job correctly will increase with time if the training has been done correctly to minimize total error.

Supervised learning generally goes through five steps. These are:

- Deciding on the datasets to use in training. This is a key step,

and if the engineers don't get this step right, the entire process is not going to work correctly.

- Gathering data. Input data for the computer must be gathered and put in the form of input vectors, with each element of the vector being a feature of the object under study.
- Choosing the structure of the algorithm to be used. There are many choices, and based on the situational requirements, the engineers will have to use their judgment for the best algorithm. Examples include regression or decision trees.
- Training. This is the exposure of the machine learning system to the training input vectors and known outputs. The training will help the system determine the algorithmic relationship between the inputs and outputs, so that future outputs can be simply calculated for new, previously unseen inputs.
- Evaluation. The engineers will evaluate the results of the training. Training with more data can be conducted if necessary. Otherwise, the system can be deployed when an acceptable level of error is reached. To evaluate training, an unrelated set of test data can be used to determine if the system can get the correct outputs with an acceptable level of accuracy.

## Issues with Supervised Learning

Supervised machine learning suffers from a problem called the *bias-variance tradeoff*. Variance is the spread - or variability - of predictions made by the system. If a system has a high level of variance, this means that it will have trouble generalizing from the training data to new datasets. In this case, it will have to overfit to the datasets it has been exposed to in training.

Bias is an error when the average output predicted by the system has a large error with respect to the correct value. Systems with high bias cannot fit the predicted outputs to the actual outputs in the training data very well.

The best approach to machine learning with systems trained using supervised learning techniques is to try to minimize the error of bias and variance *simultaneously*. You want a system with low bias and with low variance. This can be a difficult trade-off in practice.

If a system has input vectors with a dimensionality that is too small, this means that the model is too simple for the application. This will lead to high bias and low variance. The solution, in this case, is to redesign the model

such that there are more features added to the input vectors.

The opposite problem can also occur. A system might have too many features in the input vector. In this case, we have a dimensionality problem. There are too many dimensions to the problem (i.e., the number of features in the input vector, which is called the dimensionality), making the detection of patterns by the machine learning system too difficult and error prone. Systems that have too many input features are going to have high variance but low bias. The way to deal with this is for the engineers to remove irrelevant or less relevant features.

Noise is another problem that can occur in large datasets. High levels of noise can cause the system to create false relationships between inputs and outputs that don't really exist. When there is a lot of noise in the input data, engineers cannot seek exact matching between inputs and outputs. In this case, they should increase bias and seek a lower variance.

## **Unsupervised Learning**

The goal of unsupervised learning is to find previously unknown patterns in large datasets. So, in this case, the outputs are unknown for the inputs. In fact, the engineers may not even know what outputs exist for the datasets. This type of analysis is going to be used to extract previously hidden information that exists within the data.

Therefore, with unsupervised learning, pre-existing labels are not used. That is, there are no input and output labels used on the training data. It is entirely up to the machine learning system to determine what is in the data and identify it.

One type of analysis that is used with unsupervised learning is called cluster analysis. Think about the way that objects or people can be clustered or grouped together in different ways. Consider the students in high school. They are naturally clustered by class. However, they could be clustered in many ways: grouped together by height, by gender, or by height and gender. They could be clustered according to IQ or by hair color. Objects in the real world can be grouped together in an infinite number of ways, and this can be done depending on the usefulness of a given application.

When objects are clustered, the machine learning system can then extract algorithmic relationships that exist between them. This can then be used to gather insights or to act.

Unsupervised learning is also used for anomaly detection. That is, the system

seeks to find outliers and determine algorithmic relationships to detect outliers when presented with new data.

We also have *neural networks*. These are structured in an analogous way to the brains of animals. Nodes represent neurons, having weighted connections to other nodes in the systems. When exposed to data, the neural networks will increase or decrease the weightings between different nodes, learning the data in the process and developing an algorithm to use on similar data in the future, based on the weighting between the nodes.

In all cases, these algorithms are not based on any explicit computer programming or definitions of rules.

## **Semi-Supervised Learning**

It is also possible to use semi-supervised learning to train machine learning systems. In this case, the training will be divided into two distinct phases. In the first phase, the system is trained on a small subset of data using labeled inputs. That is, the first round of training is carried out using supervised learning. After this is completed, the system is then taken through another learning phase, this time using different datasets for unsupervised learning.

## **Some Current Applications of Machine Learning**

Machine learning has been rapidly deployed in the marketplace in a wide variety of applications. In this section, we are going to briefly review some of the most common uses of machine learning to date. This might help you determine ways that machine learning could be deployed in your own business operations.

- **Financial Services:** Machine learning is being used on a widespread basis in financial services. This includes fraud detection, predictive investing, credit scores, and loan approvals.
- **Spam Detection:** This is a classic example of a classification problem that has been solved using machine learning.
- **Virtual Assistants:** Siri and Alexa are examples of using machine learning to help consumers search for information and complete certain tasks using vocal commands combined with machine intelligence.
- **Predictive Analysis:** Machine learning is routinely used to give product recommendations, predict future economic behavior and conditions, and do weather forecasting.
- **Marketing and sales:** Machine learning is being used for



targeted advertising and improved marketing, helping companies identify the best marketing channels and approaches that will help them grow their businesses.

- Traffic prediction: Machine learning can be used to analyze traffic flows and redirect resources to keep things moving smoothly.
- Allocation of resources: By using machine learning, law enforcement agencies can predict where future crimes will occur and how this will vary with time. They can then allocate police forces in a proactive fashion before the criminal activity occurs. Resource allocation based on machine learning is also being used to allocate staffing in hospitals, and better positioning nursing and other staff to better meet patient needs.
- Video Surveillance: Humans can have trouble managing large amounts of video surveillance. Instead, machine learning systems can be deployed that can then alert human actors when anomalies are detected.
- Cybersecurity: Machine learning is being used to maintain the security of computer networks, prevent unauthorized access, and detect malware and other suspicious activities.
- Social media and access: Machine learning is being used for facial recognition, recognizing similar images, and gathering people you may know to help you establish better connections.
- Online customer support: By providing frontline customer support, machine learning systems free up human labor for other purposes and help businesses reduce overall costs.
- Self-Driving Cars: These are a good example of machine learning systems getting better with time by being exposed to increasing amounts of data.

## **Chapter 6: Data Science**

As big data and computer science converged both in the development of new and more powerful hardware and storage systems and with the improved development of machine learning systems and artificial intelligence, a new field that is having a large impact on business intelligence was developed. This field is commonly known as data science. Data science is an interdisciplinary subject that brings together three different fields of study. These include computer science, statistics and probability, and business acumen. All three fields lie at the intersection of business intelligence and big data. In this chapter, we are going to get a handle on the field of data science and how it's going to be used with business intelligence in the future.

### **Data and Machines in Business**

It might look obvious in hindsight, but when the Internet was invented, it increased the capability of companies to track people. No one could have predicted the way that this would merged with artificial intelligence and statistics in order to drive a new direction in technology. What the emergence of big data did was to bring together many different fields of study and their application in the real world in a new and different way.

In some ways – and for a while - machine learning was a solution looking for a problem. Big data has provided the exact type of problem that machine learning needs in order to work its magic. This marriage is part of what helps create the new field of data science. This is because a large part of the big data that has been generated is contained within the business environment. Therefore, a new need emerged for people with skill sets that would contain knowledge and abilities across multiple fields. In other words, traditionally computer scientists are not people who have business knowledge. In the old world, computer scientists would have been people whose skillset was focused on machine learning and artificial intelligence along with computer programming.

Then on the other side, you would have experts in the business. Businesspeople, including those trained formally in business school, were not the type of people that would have expertise in fields like machine learning. Therefore, they would not be able to even conceive of applying artificial intelligence to solving business problems, including incorporating it within the business intelligence structure.

In addition to all of this, there is a third subset of experts. That is the people

from the mathematics realm who are experts in statistics and probability. These people are certainly able to apply their skillsets to a wide variety of problems. But, traditionally speaking, they have not been involved in the study and application of artificially intelligent computer systems.

The advent of big data has changed all this. The first problem began when businesses were collecting myriads of data, but they had no way to do anything with it. It soon became clear that the use of machine learning was the key to solving the big data problem. After the turn of the 21st-century, machine learning began to grow in importance in the use of dealing with big data in the context of business.

This created a new set of needs. Companies that were using big data began to develop a need for expertise across fields, people who could work with machine learning and artificial intelligence applications. These specialists were not needed to make them work, but to understand how to apply them within the business. These people would also need a thorough knowledge of statistics and probability in order to work with the systems, train the systems, and understand the results that they were producing. In order to do that, the expert would have to have B/business knowledge in addition to their other skillsets.

As these things converged, the field of data science was born, and it quickly became official.

## **Comparing and Contrasting Data Science and Business Intelligence**

In the section, we need to revisit the definitions that are used for business intelligence and data science. First, let's look at what business intelligence is in relation to data. We must admit to ourselves that there is a's definition of business intelligence that is rooted in the past and that it needs to change as we keep evolving in a new situational environment. The main purpose of business intelligence has been to gather the current state of data that is behind the operation of the business for humans to make data-driven business decisions. Although recent efforts have been upgrading traditional business intelligence systems - such as data dashboards - for the most part, business intelligence is rooted in the static view of the business. In short, business intelligence looks at a frozen passed state of the business in order to move forward with new decision-making. This is a completely understandable situation, given the nature of the tools that have been available to businesses

in the past. Large databases, and spreadsheets, and so forth, are - by their very nature - rooted in a static picture that is a snapshot of a past state of the data. Therefore, the process of business intelligence is to interpret the past data that has come into the business, and most of that data (if not all that data) has been in the form of structured data. Traditionally speaking business intelligence has been used to drive reporting or the development of spreadsheets which can present data in charts and so forth, that the business can use to interpret and then make decisions about future actions. This type of data has not been used to make predictive decisions for the most part.

So now let's look at what data science is, and how it operates. What we have done so far is to define the major components of data science that are used as tools. Those are machine learning, or artificial intelligence, business knowledge, and statistics. But these are tools, and they should not be confused with data science itself. Data science, like business intelligence, is a data-driven enterprise. However, data science looks to extract hidden patterns, meanings, and insights, from structured *and* unstructured data. One of the key differences between data science and business intelligence is that data science seeks to either predict future outcomes or to be able to take advantage of unknown patterns in future data.

The naïve observer would look at this and believe that business intelligence and data science are completely different. They seem to have different goals, and they're applied to different situations. They do have some things in common on the surface. That is, both data science and business intelligence are looking to improve the competitive advantages of the enterprise. This can be done across the entire operational procedures of the business. It can impact the allocation of resources, customer service, expansion into markets, increasing efficiency and avoiding wasteful spending, and other things that impact the general efficiency and success of the business. Thus, everyone will agree that to a certain extent, data science and business intelligence can have the same end goals for the company in mind.

However, one should step back and look at what business intelligence really is. The fact is, business intelligence is going to grow and evolve in this new environment, and it will incorporate data science into the business intelligence model and process and use it to help drive not only increasing efficiency, but also actionable decision-making. It is not necessary for upper-level management and senior executives to have detailed understandings about the workings of data science. It is up to the data science team to extract

useful information from their activities that can be used for important decision-making by upper-level executives.

It is also a myth that business intelligence must be constrained to using structured data. Data science represents an opportunity to expand business intelligence into a broader framework that can utilize the huge amounts of unstructured data that companies are collecting both externally and internally. We have already described the ways that unstructured information that a company has access to could be used to improve the efficiency of the company massively. Think about what some of the end goals are business intelligence when it comes to using the data that it does use in a traditional sense. The problem that most analysts have had when it comes to looking at business intelligence is there to focus on the sources of the data in the way that they are presented and utilized by the business. Instead, you should focus on the end results that you would seek to achieve through the application of business intelligence. When you do this, it is easier to adopt a broader horizon within which you can incorporate the tools and sources of data used in data science.

Let's take a concrete example from UPS. It is well known that UPS utilized Machine learning systems in order to identify patterns exhibited by their drivers during day-to-day operations. From this effort, the company was able to identify changes that could be made to reduce fuel consumption across the company. These changes were then implemented. So, if we ignore the source of the data, we can see that this is the kind of problem that we might address using business intelligence, if only we had the information available. The only difference here is one of technology. That is in a traditional sense; the type of data needed to solve this problem simply was not available using the tools that have been traditionally in existence when it comes to Business intelligence. But you can read any business intelligence book or article and find copious examples of how one of the goals of business intelligence is to improve operations in the business — furthermore, decisions made using business intelligence our data-driven decisions. So, looking at the UPS example, where is the distinction between what the data is telling you and then making a data-driven decision based on it, or saying it is data science or business intelligence?

The reality here is that we are looking out data science in the wrong way. Instead of seeing data science as a distinct entity, we should instead view data science as an additional tool that can be used as part of the overall process of

business intelligence. Too many people who are experts in business intelligence are making these mistakes because they are uneducated or lack proper education when it comes to data science. Quite frankly, the understanding of many business intelligence experts of data science is confined to a top-level understanding of it that they glean from reading articles and publications. In fact, we can go so far as to say that some people in the business intelligence community are threatened by data science.

However, when you step back and look at how data science can be used, it's clear that these fears are completely misplaced. There is absolutely no reason for the business intelligence community to try to keep data science outside of it or to even fear it. Instead, it should be embraced and incorporated into the totality of business intelligence.

You will read articles about business intelligence that claim that it can only understand data that is structured and formatted. That is completely absurd. It is limited thinking that is only viewing business intelligence through the framework of current technologies. Buy an analogy, let's imagine that it's 1975, and the issue is the introduction of electronic computers together with word processors and spreadsheets. If we were using the logic that some in the business intelligence community are using now, you would have said that spreadsheets could play no role in the accounting department, because accountants are working with pencil and paper ledgers. But what happened as spreadsheets were adopted? The accounting process changed and incorporated spreadsheets within it.

This is what is going to happen when it comes to the relationship between business intelligence and data science. The reason that this is going to happen is that data science and business intelligence have the same goals when it comes to meeting the needs of improving the competitive advantages of the business. Furthermore, both data science and business intelligence seek to promote data-based decision-making.

Also, you should not get caught up in the situation of focusing on the fact that in many cases, artificial intelligence systems are able to work autonomously. That is another side issue that only serves as a distraction. Again, I refer you back to the earlier two examples provided (UPS and Southwest Airlines), which were able to make very important decisions that improved the efficiencies of their businesses while saving hundreds of millions of dollars. In both cases, this was done using machine learning and data science. It is clear to this author that looking at the results they could be considered to be

something that would fit within the overarching realm of business intelligence if you're willing to expand it and update the notion of what business intelligence is going forward into the future.

## **Improved Efficiency Through Automation**

Remember the V's that we talked about regarding big data? One of the V's was volume, and another was velocity. These two are going to increase exponentially as the years and decades pass. That means businesses are going to be forced into a situation where they need to rely on data science and data scientists in order to interpret and understand the big data that is playing a role in their business.

The intern is going to be increasingly influenced by artificial intelligence and autonomous computer systems. This is only going to serve to increase the efficiency of business intelligence massively. Many aspects of business intelligence that are traditionally done by humans are going to be increasingly automated and made far more efficient. That is going to magnify the power of business intelligence exponentially. It is going to be an incumbent on business intelligence experts to bring themselves up to speed on data science; they will have to increase their understanding of statistics and machine intelligence.

This requirement is often misunderstood and often framed in terms of an either/or situation. That is, we are returning to the view that business intelligence is somehow threatened by data science. But once again, this is a complete misunderstanding. Rather than looking at data science as being competitive to business intelligence and making excuses about people in business intelligence not understanding or being able to understand the workings of data science, we need to be realistic about this. Does someone in business intelligence understand all the computer code that is behind a database, or the internal workings of spreadsheet programs like Microsoft Excel?

Of course not. What they do understand is the high-level aspects of these tools. So, when it comes to incorporating data science into business intelligence, a high-level understanding of the tools is what is necessary. In this way, data science is going to enhance the ability of those in business intelligence to make it effective and actionable decisions for the business. Once again, we have already seen how this works in practice. Data science is helping companies improve their efficiency, break into new markets, allocate

resources effectively, massively expand, and improve their marketing, advertising, customer service, and much more. Let's look at some potential applications a big data and data science that can have exciting uses within the context of better business intelligence.

Until recently, and even in the present for most companies, huge amounts of useful data that can improve internal operations are simply sitting collecting dust on hard drives. Companies need to analyze ways that employees are communicating with each other and using their time in order to improve efficiency.

We are starting to see this with monitoring systems based on artificial intelligence. Companies are so much as snooping, because you are working on company time, and the organization has every right to see how their employees are using their time and resources. For those companies that have started using these tools, increasing the level of productivity and efficiency among employees has been one of the many benefits to come out of this. It makes it easier to weed out bad employees and makes it possible to bring problems to the attention of good employees, allowing those employees to rectify the situation. By studying the movements of employees, it's also possible to reallocate the positioning of employees in the company to increase the efficiency of collaboration between different groups and employees. This is analogous to views of artificial intelligence in hospital environments wherein nursing staff has been completely redesigned to consider the needs of patients throughout the facilities. So, in the same way, a company that really doesn't understand why it distributes employees among offices the way it does can instead put employees in locations where they are going to be needed most of the time. Until now, the tenancy has been to distribute employees in the building based on the group or department they belong in, considering where any free office space exists. What if instead, you actually studied the data derived from tracking employee movements where they were spending most of their time that was devoted to productive activity and then you brought employees closer together that actually needed to be close together in order to get work done?

## **Self-Service BI**

Earlier, we mentioned that one of the capabilities of artificial intelligence systems is that they can work autonomously without human input. One of the developments that are coming out of this is the idea of self-service



business intelligence platforms. This will allow companies to increase efficiency as far as incorporating business intelligence throughout the organization. It will be easier for those who need business intelligence services to get it exactly when they need it. Furthermore, it can help companies become more efficient by reducing staffing requirements.

In addition, artificial intelligence systems can learn, and they will continue to learn and become better at what they do the more they are used. Every time an artificially intelligent or self-service business intelligence system is used, the machine is learning and improving itself in response to new data that is coming in. What does this mean for the business? It means that it will have business intelligence that is consistently improving itself autonomously. Furthermore, artificial intelligence systems can work 24/7 without becoming tired or complaining. As time goes on, this is going to mean that self-service business intelligence platforms are going to grow in power exponentially. This is only going to help the business skyrocket its efficiency and productivity.

## **Labor Saving**

Another benefit of incorporating machine learning and business intelligence is that just like using machines and other contexts, data science is going to free up human labor that can then be used for other purposes. Unfortunately, far too many people are completely misreading this situation. Instead of looking at it as the opportunity that it is, they are viewing it with fear. It is rather disconcerting to see large numbers of smart people who giving in to the old Luddite philosophy that has been proven wrong repeatedly throughout history. For some reason, these people are viewing the tools of data science and artificial intelligence as somehow different than other machines, including computers applied in different circumstances.

What's important to understand about this is the fact that the nature of the machine being used is not the relevant factor here. This is because when human beings are freed from labor, they invent entirely new things to do, things that they didn't imagine anyone would want to do *before the changes came into existence*. There are many examples of this that have taken place since the development of computers. For example, one of the most popular activities on iPhones and iPads is playing solitaire. If you were able to travel back in time and ask someone in 1960 whether people would want to play solitaire or poker on the phone, they would have looked at you as if you were

completely crazy!

Now consider one of these side effects that is occurred as a result of people playing games like this on their phones. This has created a massive business that didn't exist before 2008, devoted entirely to developing games that can be played on mobile phones. This industry is developing so rapidly and growing so fast that in just ten years it has grown into a multibillion-dollar industry that employs hundreds of thousands of people directly, and hundreds of thousands of additional contractors.

Even if you went back to the year 2005, such a scenario might've been deemed impossible.

The point of the story is that new technology leads to unexpected developments that increase the employability of people. Yes, they need to learn new skills. Before 2008, there probably weren't many people that knew how to create a card game on a mobile phone.

Likewise, as business intelligence becomes more automated (and to a certain extent more dominated by data science<sup>0</sup>, that doesn't mean that the people in business intelligence are going to have to go on unemployment. What it is likely to mean is that businesses are going to find new and unexpected ways that business intelligence can and will be used in order to improve the performance and competitive advantages of the business. Therefore, my prediction is that ten years from now there are going to be more people employed in business intelligence than there are now.

Let's look at one of the final benefits of incorporating data science into business intelligence. By automating more tasks that are associated with business intelligence and increasing the role of unstructured data, this is going to free up the enterprise so that it can focus on goals and outcomes rather than on tedious tasks and efforts to interpret data. Data interpretation is going to be made far more efficient and productive as a result of these improvements.

So, this leads one to believe that data science is going to help massively improve the outputs that result from the use of business intelligence while also increasing the efficiency of business intelligence. Rather than looking at data science and business intelligence as being in conflict or at best as having an application into distinct realms, we should instead focus on the fact that these two fields are going to merge and be complementary. Later, as the tools and results of data science begin to seem natural, you might even start viewing data science and business intelligence as being two sides of the same

coin.

When it comes to developing and maintaining a competitive advantage for your business, there is no question in my mind that you're going to need to incorporate both data science and business intelligence together in order to survive the future hyper-competitive environment. If you are not doing so, I can guarantee that your competitors *will* be doing it.

## **Chapter 7: Cybersecurity**

At the foundation of business intelligence, we find data. This data is in many forms and may be stored in many different locations. We have seen that businesses include a lot of structured data that may be stored internally on their own computer servers. It is accessed in the form of relational databases, spreadsheets, and other data structures. Companies are also collecting and storing a lot of unstructured, internal data as well as data from customers and external sources that may have unstructured formats. All this data represents key components of the business operation. It is important for the business to maintain the security and integrity of this data. As a result, cybersecurity is going to be playing an increasing role in the future of business intelligence. If the data that a company has is not secure, then business intelligence is not going to be operating at the level it should.

### **Determining Infosec Needs of the Company**

Cybersecurity is going to be involved with business intelligence in multiple ways. One of the ways that it's going to be involved is that upper-level management and security teams are always going to need to be able to evaluate the security of their data. This is one area where Machine intelligence and data science can play a central role. Machine intelligence can be used not only to secure data but also to identify and present security weaknesses that exist in the organization. It will be up to the management teams to determine what to do with this information. However, there's no question that using the capabilities of machine learning weaknesses that can be exploited and other to threaten the company and the security of its data is going to be identified in unexpected and new ways. This is only going to serve to improve and enhance the ability of the company to respond to information security threats as they arise.

So, what we are envisioning in this context is using machine learning in order to identify threats that exist for the information that is so vital to the company's competitive advantages. This fits in well with business intelligence for several reasons. The most significant reason is that it will enable involved management concerned with information security to make data-driven actionable decisions. Of course, these decisions are going to be involved with the usual kinds of trade-offs that businesses always face, such as practicality considerations and cost-effectiveness.

## Cybersecurity in a nutshell

At this point, let's look at what cybersecurity involves. In today's world, many of the key components of the business include software applications, computer systems, computer networks, and data. These systems are in constant threat from malicious actors throughout the world. One of the side effects of the internet that we may not like is that malicious actors, located anywhere, are able to locate, detect, an attack any computer system that they choose. An interesting aspect of these malicious actors is that although a lot of the news is focused on large-scale hacking attacks, it is often small businesses that are the targets of these kinds of efforts.

Let's look at some of the types of attacks and how they occur. The most common that people focus on is a hacking attack which breaks into a computer system for the purpose of stealing data. This is a classic case of breaking into a large business computer system in order to get the names and Social Security numbers of people that hold credit cards.

However, there are many other types of attacks that can occur. For example, you can imagine that some unethical people would like to deal with competitive threats not by stealing people's credit card numbers, but instead by tampering with the data that a company has. Throughout the book, we have discussed the importance of data that is becoming critical for company operations. Think about how you could damage a company by breaking into their computer network and not stealing anything, but rather tampering with the data. If this attack went undetected, that data would be used for all kinds of purposes. All the machine learning and analysis that we've been talking about would be continuing based on tampered data that might even be fake. The company would be led down the road to making bad decisions based on bad data.

When it comes to small-scale attacks, ransomware is a common threat. This is used to extract - or at least - attempt to get money from victims.

As time goes on, all the above attacks and more are becoming increasingly sophisticated. There's basically an arms race going on between the cybersecurity community and malicious actors worldwide. Interestingly, many of these malicious actors are not individuals or teenagers working out of their mother's basement. In fact, many countries such as China, Iran, Russia, and North Korea, have official government programs that promote hacking attempts and cyber espionage. Considering that in some of these countries - especially China - many corporations are state-owned or partially

state-owned, it doesn't seem far-fetched to imagine that they would not be above tampering with data of competitors in other countries.

## **The Approach to Cybersecurity**

Modern approaches to cybersecurity are based on a five-step process. This includes identifying threats, protection from threats, detection of threats, repeat and refine.

Identifying and detecting threats is increasingly coming under the domain of machine learning. This was one of the earliest applications of machine learning which had a practical use. By using anomaly detection and classification algorithms, machine learning systems can detect attempts to break into a network that are probably malicious in intent. Many computer systems are under constant attack and the ability of machine learning systems to adapt while spotting anomalous behavior makes them well-suited to playing a central role in cybersecurity.

## **Data Science and Cybersecurity**

Big data has a large role to play when it comes to cybersecurity. By utilizing big data, it is possible to find trends in the underlying data related to cybersecurity. This can be used to determine patterns of attacks that are made. It is impossible to know what those patterns are going to be ahead of time; however, you can look for things such as attacks during certain time frames. Once that is determined, it can turn from a raw piece of data into actionable Intel.

Data science can only help with this process. One of the ways that it can assist with an actionable intel is that it's going to be possible to use unstructured machine learning in order to detect unexpected patterns, relationships, trends, that exist in relation to attacks on computer networks and systems. For example, a machine learning system may detect the most likely times of day, departments, or even computer terminals that are most likely to be used in an attack. When this information is known to the security team, this is data that can be used in order to make actionable business decisions to increase the security of the company.

Another way the data science can be used together with machine learning is to develop more secure systems. That is rather than having humans develop new login systems or encryption, you can put the task to artificially intelligent systems. They are more likely to develop protocols that are hard to break and add more security. One of the ways that they're going to be able

to do this is that you can train the machine learning systems by giving them data over the past history of the organization and the computer networks that the system can learn from in order to detect patterns of attack and also understand the types of attack that the system is most likely to encounter.

Big data is going to relate to cybersecurity in two different ways. We have already determined that big data can be used to train machine learning systems to understand the patterns and past practices that have been used by malicious actors in order to try and break into the computer networks. However, big data is also a liability. For many companies, it is the presence of all the data they hold in the first place – that cache of information is making them subject to attack. If we are going to say that data is the gold of the 21st-century, it makes perfect sense that people the world over are going to try in access that data in malicious and illegal ways. So, the more data of the company has, the more important cybersecurity is going to be to the organization.

By utilizing big data and machine intelligence and cybersecurity, businesses will be able to develop more accurate and effective defense systems against attack on their computer networks. In addition, this type of information can be incorporated into business intelligence and help management develop more cost-effective solutions. In other words, management will be able to maximize cybersecurity while keeping budgets at the most efficient levels possible. They will be able to precisely target where money needs to go in order to maximize cybersecurity going into the future year.

## **The Three Tenets of Cybersecurity**

It is often said that cybersecurity is based on three tenets that are called the “CIA triad”. Maybe the name is a little bit unfortunate. In any case, C stands for *confidentiality*. Confidentiality measures are used within cybersecurity to protect data from unauthorized access. This is one area where clear application the big data together with machine learning can be used to make confidentiality protections better.

The second area of cybersecurity is *integrity*. We have mentioned this earlier. There is a good reason for malicious actors, especially if they come from other companies or nationally directed enterprises to tamper with data rather than expose it. Integrity is the component of cybersecurity that would protect data from unauthorized changes. The types of changes that could be used include changing the data itself. That kind of attack would ripple

throughout the enterprise. So, this is a very serious part of cybersecurity. Data also must be protected from unauthorized access that would add new data up to the datasets or delete data.

The third tenet of cybersecurity is *availability*. It is important for cybersecurity to function in a way that is basically running in the background. It can't interfere with the normal operations of the company. Users in the business are always going to need to access data practically already keep the company running at peak efficiency. Therefore, availability is going to be important for those in cybersecurity to consider.

## **Threat Detection**

Of course, one of the most important aspects of cybersecurity is going to be threat detection when there are attacks on the network and attempts at unauthorized access. In order to work best, this type of threat detection must be operating in real-time. This is not something that humans are well-suited to do. You can immediately draw the conclusion that this is a task that would be best propose for artificially intelligent systems. In fact, you won't be surprised to learn that machine learning is applied to this very task all the time. Threat detection is a simple task for machine learning systems, if you have the data necessary for them to learn on. By presenting machine learning systems with past data of attacks on computer networks, they're able to learn and spot anomalous behavior that would indicate an unauthorized access attempt of the network. Businesses and organizations may be, to some extent, better off if they do this type of analysis on their own if they can do so. The reason is that attacks directed at a company may be unique in nature to a certain extent. For small businesses, they may have to rely on off-the-shelf software products. These products will still be very effective because they have been trained in a wide range of possible cybersecurity threats.



## Chapter 8: BI and Data Mining

The core idea behind business intelligence is to use the data that a business has available in order to develop actionable information. This will help the business operate more efficiently because management will be able to make data-driven decisions rather than trying to act on incomplete information and hunches. As it stands on its own, data is not useful in its raw form. This is where business intelligence comes in. It is going to take that data and make it in a form that can be presentable to two human beings that can then make informed decisions based on what the data is telling them.

As such, business intelligence will pull together all the data in the organization, analyze the data, present that data in report form or in a visualized way, which will make the data meaningful for management. Then in order to enhance the competitiveness and efficiency of the business, management can make data-driven decisions.

### Data Mining

Now let's become more acquainted with data mining. Data mining is a jargon word in a sense. It already has a lot in common with some of the things we've been discussing. The first thing that data mining is involved in is *large datasets*. In

other words, here we have big data yet again – but that is only in first appearances. In fact, part of data mining is “mining” the data, finding smaller subsets within the large datasets that are useful for the analytical purposes at hand.

Another thing the data mining is involved in is recognizing hidden patterns that exist in these large datasets. Thus, we are back to the tasks that are carried out with machine learning, although this isn't explicitly specified when discussing data mining. Data mining attempts to classify and categorize data so that it's more useful to the organization.

So, we start with raw data which is basically useless. Data mining helps convert that data into something that can provide value as far as the information that it contains. A part of data mining is going to be selecting the data that you want to use. Data warehousing is an important foundation upon which data mining is based. Companies need to be able to store and access data in large amounts which is why data warehousing with effective solutions that are fast and accurate is important. Then the data must go through a cleansing process. That is when you have huge amounts of data, one of the

problems that you're going to encounter is that data is going to be often corrupted or missing. This is something that is very common when it comes to relational databases, but this can also happen when you're restoring huge amounts of unstructured data.

After the data has been gathered, extracted, and cleansed, the process of data mining moves on to look for the patterns needed to gather useful information from the data. Once this is done, the data can be used in many ways by a business. For example, it could be used for sales analysis or for customer management and service. Data mining has also been used for fraud detection. There is much of overlap between data mining and other activities involving big data, such as machine learning. When it comes to data mining, you're going to see a lot of statistical analysis.

This intelligence and data mining are both involved in the process of converting raw data into actionable information for the business. However, the goal of business intelligence is to present data in meaningful ways so that management can make data-driven decisions. In contrast, data mining is used to find solutions to existing problems.

If you remember when we talked about big data, one of the things that were important was volume. Business intelligence is certainly driven by large datasets. However, data mining is different in this respect. Relevant data is going to be extracted from the raw data to be used in data mining. Therefore, relatively speaking, data mining is going to be working with smaller subsets of the data that is available. This is one characteristic that is going to separate data mining from the other topics that we have talked about so far. Data mining might be used as a part of an overall strategy of business intelligence. So, what management is looking for from data mining is solutions that can be applied to business intelligence. This contrasts with business intelligence on its own, as it is usually used to present data to people.

So, the core result obtained from data mining is knowledge. This is in the form of a solution that can be applied within business intelligence. This provides a big advantage to business and operations. That is because the findings from data mining can be applied rapidly within business intelligence. Data mining is also a tool within business intelligence that allows business intelligence to extract complex data, presenting it in understandable forms that are useful for the people in the organization. The data extracted with data mining can be presented in readable reports or in graphical format containing graphs and charts. In this form, it becomes a part of business intelligence so

that the people in the organization can understand, better interpreting the data and making actionable decisions based on that data.

The volume of data coming to large businesses is only growing with time. This makes both data mining and business intelligence more important to the organization as the onslaught of information continues to pour in. It is going to be important to cull the data in terms of saliency; this is where data mining plays a role. The data is always changing, making this task even more important. Demand for data mining and business intelligence solutions will be increased in proportion to the growth of the volume of data.

For companies to remain competitive and - especially if they want to be a market leader - they are going to have to utilize data mining and business intelligence solutions for retaining their advantages.

## **Data Analytics**

Data is not useful if you cannot draw conclusions from it. Data analytics is a process of organizing and examining datasets for the purpose of extracting useful and actionable information from the data. Data analytics plays a role in business intelligence, using tools like OLAP for reporting and analytical processing. When done effectively, data analytics can help a business become more competitive and efficient, build better and more targeted marketing campaigns, improve customer service, and meet the goals that are a part of business intelligence. Data analytics can be applied to any data that an organization has access to, including internal and external sources of data. It can use old data or even real-time data to provide more readable information that can be accessed by employees in the organization in an effective way to help them make actionable decisions.

While data analytics can be used as a part of business intelligence efforts, like machine learning, data analytics can be used for *predictive modeling*, which is not part of business intelligence. Typically, BI is used for an informed decision-making process based on analytics of past data. Data analytics uses past data but can apply it with predictive analytics to help the company use modeling and tools to determine future directions of various efforts that can help the company maintain its edge and advance even further.

Data analytics will also be used in many ways that are like processing data with machine learning. That is, it will be useful for pattern recognition, prediction, and cluster analysis. Data analytics is also an important part of the data mining process.

## Chapter 9: BI and Social Media

Over the past year, the data collection powers of the social media companies have come to the forefront of many discussions. At the top of the list of concerns is privacy. Regardless of what you think about these discussions, one thing is clear: social media has resulted in the collection of unprecedented amounts of data - not only about individual people but also about businesses that use these platforms. Social media is a very effective way to collect data on customers.

Social media represents unprecedented opportunities for businesses. For one thing, social media will help businesses understand the behavior of their customers. Social media also helps businesses target the market to new customers and acquire them. It also provides an opportunity for a business to put their face forward in new ways.

In this chapter, we are going to look at the power of social media in terms of interaction between business intelligence and social media; we will explore how that can help businesses expand and improve their competitive advantage.

### Leveraging Social Media

There are many ways that a business can leverage social media. The first way is to recognize that companies like Facebook and Twitter have an imaginable treasure trove of data on every customer. The data collected is thorough and global. Moreover, companies like Facebook are ahead in terms of organizing that data and putting it into a useful form. So, the existence of social media companies not only provides a platform through which a business can increase awareness, but it has also created an environment where other companies are doing a lot of the hard work for you. So, there's a bonanza cache of unstructured data that is not only being stored by companies like Facebook, it is also being analyzed. So, let's begin at the beginning: the first advantage that we have here is that we don't have to worry about storage capacity because Facebook or Google already has that data stored for us.

This data has also been put into a form that makes it friendly for all kinds of analysis. And although there is a lot of hype about privacy violations, the facts show that for the most part, data is presented in aggregate ways, preventing the targeting of any individuals *unless those individuals voluntarily choose to interact with other companies*. "Voluntarily choose"

means that an individual has freely given their name and email address to the company for its use – and has read, understood, and agreed to the company’s privacy policies.

## **Web Scraping Tools**

Web scraping tools allow a company to obtain data that is on other websites (publicly available data) without having to copy and paste it manually. These types of tools allow businesses to get data from social media sites that can then be analyzed and used. This data will be in an unstructured form, and as such, will be well suited for analysis using machine learning. No matter which social media platform you are scraping data from, you are going to have mixed data in almost every case. Consider a Facebook post as an example. A Facebook post may have an image associated with it, but it might be plain text. It could also include a video. It might have a hyperlink and emojis. So, there is a mixture of data that is contained in one single object. This is also true on many other social media platforms; while Pinterest and Instagram are photo platforms (primarily), postings will have text, hashtags, and possibly hyperlinks.

In order to get usable information with this data, it must be rigorously “percolated” via a data grinder. As we mentioned, this is clearly big data, and it’s also unstructured data. That means it is not particularly well suited for use in traditional business intelligence, but rather must be processed using big data and machine learning methods. You will also be searching for hidden patterns in the data. As an example, if you were looking for a hyperlink appearing in many Facebook posts, you would need to cross-link it with the people that are posting the link in order to find their demographics.

Web scraping might seem like an insurmountable task, but it can have many advantages. It provides a way to collect data that can be used for marketing research. It can also be used to extract contact information. However, the value of doing that is questionable, as most people don’t respond well to contacts that they – themselves – have not initiated.

There are many effective web scraping tools that are available. You can consider using import.io, dexi.io, and visual scraper. These powerful tools can help you do some pre-processing on the data, assisting you in retrieving the desired type of data that are useful for your purposes. In some cases, you might be able to get it into a form that can be directly used in business intelligence.

## Direct Interaction with Customers

One of the interesting benefits of social media is that it provides businesses with the ability to interact with customers directly. This is going to involve some effort on the part of the business; the more effort that is put in, the more likely it is going to pay off. And the fact is, this is easy to maintain, and can be done on a low budget. In the following chapter, we will be discussing the types of advertising that can take your direct interaction with customers to a new level.

The first step toward direct interaction with customers is to create a Facebook page for the business. A Facebook page is essentially like a Facebook profile, but they are created either for group interests, hobbies, or in the case of interest to us businesses. The Facebook page is going to have a timeline and a photo repository, just like any normal Facebook profile would. One of the mistakes that a lot of businesses are making: they create Facebook pages, but then fail to maintain them. If you're a small business, it's vitally important that you maintain the Facebook page. It doesn't take a lot of work, and one or two posts per day are enough to grow the Facebook page over time.

Unfortunately, what you often see when visiting Facebook pages for businesses is very few posts at all, and many just create the page and leave it there as a placeholder. Among those that do post, they don't do so effectively. Your post must be engaging, inviting users to comment on and share the post. As we will learn in the next chapter, we can use advertising to bring in more users.

But the main benefits of the Facebook page are that once somebody likes the page, your posts will start showing up on *their timeline*! This is only one of the many benefits of marketing by posting regularly on a Facebook page: when your posts are shared, it creates a domino effect! Friends of the person who originally shared will see your post, and they may share, and then - you get the picture! By posting interesting, engaging content, you are reaching far more people than just those who happen to see your post in the first place.

A side point to note: user comments left by viewers on your Facebook page are data. Even if you are running a small-scale operation, these comments can be analyzed by those at your company, providing actionable information.

## The Challenges of Social Media

As a business, one of the challenges of social media is determining which platforms are the most useful for the purposes of your business. The first factor to consider is the question: what are your main customer demographics? There are certain social media platforms that are used more frequently by young people, whereas other social media platforms are used by the general public. The form of data on social media platforms may also vary from platform to platform.

Let's get started by looking at a few examples. One of the most popular social media platforms (and one that does not get much press) is Pinterest. This platform has been around for a while, and although it has a mobile app now, its original introduction to the public was as a website. This website's purpose is to share user images, and its primary audience is female. Of course, that doesn't mean that males are not on the website; however, in proportional terms, the audience tends to be female, and more specifically, it tends to be females of a certain age group. It is estimated that the age group of 18-30 makes up most active users on Pinterest.

Let's contrast this information with that of Facebook and its users. Facebook has been around for a long time, becoming dominant around ten years ago. The advantage of Facebook is that most people already use it in a personal manner, so the audience is already vast. If you are targeting Facebook and looking for the advantages, the biggest one is already baked in: you are going to be able to reach nearly every demographic that there is! If you are using social media to leverage people over the age of 50, Facebook works for you. It would also work perfectly fine if your customer demographic was 18 to 34. Now let's look at Instagram, which is owned by Facebook. Instagram is an app-only interface. This fact alone probably makes it more appealing to younger people. Although Instagram has recently become used more by older people, it remains primarily the 18 to 34 age group. Gender is more balanced on Instagram as compared to Pinterest.

Twitter is another platform that, like Facebook, seems to appeal to all age groups. One downside of Twitter is that due to the nature of the platform's communication protocol, advertising on Twitter is a bit more difficult. Some businesses may have trouble connecting with users due to this flaw. That said, it has 600 million active users, and so it can *reach a lot of people*.

There are other social media platforms such as Snapchat and WhatsApp. These are confined to the mobile space and appeal mostly to people under the age of 30.

This is an incomplete review but helps to demonstrate that you need to choose your social media platforms carefully based on what your business is doing, as well as its demographics.

## **Data Issues with Social Media**

Social media can present data in a wide variety of forms. One of the first factors that you need to consider is heterogeneity. This means we need to examine the data in terms of the *data types* included in social media postings. So, if heterogeneity is strong, it means that the data is taking many different forms. Think multimedia here. Data may be in the form of text, images, and video – or other types that you don't readily consider but may be extremely useful. For example, the hyperlinks that people put in their social media postings are going to tell you a lot about that person. Hashtags are also important to look at. As you can see, social media postings contain more data than meets the eye at first. By looking at hashtags, it may be possible to glean information such as political affiliation for things that the poster is interested in.

The viral nature of social media is also something to consider. One question you must ask: is a post by an individual an *original post*, or is it a *shared post*? Even if a person shares a post, it may still provide data about that person. More than likely, if the person shares a post, they either found it interesting or they share the views or interests that are present in the post. They may just think it's funny. These factors can present a challenge to businesses trying to do an analysis of social media data.

It is possible to retrieve data from social media sites using web scraping tools. That takes a lot of processing power, and how it handles and separates that data presents a difficult choice. Data will have to be sorted by source, and by type. Different types of analysis can be applied to the data once you have an in your possession. Clustering analysis could be very helpful. For example, you could consider looking at all the posts that share a particular link. Then an analysis could be used to determine the characteristics or other data points of the people that share the link. This can be important for marketing purposes, and it's even used by political professionals that are trying to target people that may be open to their messages.

Another issue related to social media data concerns immediacy. If you're going to utilize social media data, you are going to need to know how recent the data is. Data from five or ten years ago probably isn't going to be very



relevant. The more immediate the data is, the more valuable the data is. Second, social media provides huge amounts of data. This is the scaling problem. For organizations working with this data on their own, this can be an extremely difficult problem to deal with. This may force organizations to either: 1) have companies like Facebook do analysis on their behalf (probably the most efficient way to do it,) or 2) they can work with smaller datasets and try to do it themselves. That process, however, it's probably not as effective as it would be utilizing the power of the social media companies themselves in order to help your business.

## **What Can You Get from Social Media?**

Let's take a step back and think about what we can get from social media. Social media offers unprecedented power that we can use to investigate the private lives of people. This is not to say that you should be snooping on people. We are talking about information that these people have voluntarily chosen to make publicly available. This information is extremely useful from a business perspective, allowing you to determine many things about different people, including interests, hobbies, and goals. People often list important information with their profile that can be combined with information such as age, gender, and birthplace. All this information can be combined and analyzed using machine learning capabilities in order to extract useful patterns for marketing purposes.

One of the keys to social media is you want to be able to speak directly to your consumers in ways that they can relate to. By analyzing data that is either shared by the social media company for you (or that you've scraped,) this will help you connect better with your customers.

## **Dashboarding the Data**

One way that many businesses deal with data from social media companies is by using dashboards. So, for example, they can ask questions about people with different demographics. This information can be gathered from social media sites. In fact, Facebook provides a dashboard-like interface that you can use to analyze Facebook users in terms of many characteristics such as education, websites they have shown interest in, and more.

For larger organizations, it may possible to integrate social media data and develop internal data dashboards that can be used in order to access and analyze the data. For example, you can create clusters of users by various characteristics. You might be interested in males, ages 30 to 44, unmarried.

Then you could use the data dashboard to extract different information about this demographic. This is all part of business intelligence.

Hopefully, you can see where this is going: the result is that we can create actionable information from these types of analyses.

## **Three Ways to Use Social Media**

The businesses that have the most success with social media are those that understand the central premise of social media. Take a step back and think about what Facebook is about. The central theme of Facebook's connection. Put differently, social media is about building relationships. This means that there are going to be three ways that a business can utilize and exploit social media.

The first way is to build relationships with your client base. This can include active clients as well as prospective ones. That is one reason why you should create and maintain a Facebook page and actively engage with the people that post there. This will go a long way towards getting your company to build relationships with the customers. You want to make it personal and real. People have a certain type of radar that they can use on a subconscious level to detect whether something is genuine or not. This is not to say that people are not possible to fool; of course, they are - Bernie Madoff proved that! Even so, people do have a certain sense of these sorts of things. Businesses should go into social media with the intent of being genuine and building real relationships with people. Even if a company is a large organization, it would benefit by putting an actual face to the page. So, it might be possible to have an employee who is dedicated to running Facebook pages and interacting with customers.

The second way that social media can be utilized includes what we have been discussing up to this point: data gathering. This is the first step to incorporate social media data into your business intelligence. From there, they can be analyzed using the tools that you already have. The applications of using this data are going to depend on your business needs and the reasons that you're collecting this data. It might be to get a picture of the ideal customer. This is a very powerful way of marketing if you can get an idea of who the ideal customer is. Alternatively, you may be breaking things down, using clustering in order to determine what different groups are driven by in terms of their interests and desires. This will help you market to different groups using targeted methods that are going to speak more directly to them.

In summary, social media can help you develop more effective relationships with consumers and more effective marketing tools. The way that this information should be used is to communicate directly to users in a way that touches them directly and strikes a sensitive chord in them.

Third, you are going to want to use social media to put your advertising and marketing efforts on steroids. This is a topic that we are going to discuss in the next chapter.

## **Chapter 10: BI and Internet Marketing**

When it comes to marketing and business intelligence, there has never been a more powerful era for companies. The seemingly endless data streams that are available for businesses allow them to market in unprecedented ways. People have been noticing this, watching artificially intelligent driven systems customize advertising for each individual person using the Internet. Many people find this annoying, but the truth is that it *works very well for business*.

Even if you're running a small business, there are many tools that you can leverage in order to elevate your internet marketing efforts. Facebook has moved to the forefront of Internet marketing and is a very powerful player in the space. But you should not limit your approaches to advertising and marketing online to Facebook alone. In this chapter, we are going to explore issues around internet marketing and business intelligence.

### **Social Media and Internet Platforms Best Suited for Online Marketing**

Let's begin by giving an overview of some of the social media and Internet platforms that are best suited to effective Internet marketing. When it comes to internet marketing, given all the data that is available, you want to think in terms of laser targeting and focus. In fact, one of the mistakes that some advertisers are making is there advertising the old way. What do we mean by that? The old way to market was to simply throw things out there and hope that the right people see it and act based on the advertisement. With all the data that is available today, that is a bad approach to take to marketing.

Let's look at a specific example. You can browse general websites, and they often have a lot of advertisements on them. Consider CNN as one example. You might think that advertising on CNN is a great thing to do. After all, they have huge numbers of eyeballs on the website all the time; people are constantly checking for the latest news. However, unless you have a business that is oriented toward the general public at large, advertising on a website like that is going to be a waste of money and time. The reason is that a small fraction of the audience is going to be interested in your product or service. For the most part, you're going to be spending precious dollars to put an ad in front of people that don't care about your services or product. In fact, most of them are never going to care about your company! Why would you want

to spend dollars doing that? Spending money on untargeted advertising is a poor choice when you can use other services to put your ads right in front of the people who want your product!

However, there are exceptions to this. Many times, after you have looked at a product on one web page, you'll soon discover that suddenly ads are popping up for that product on other websites you browse. For instance, you are considering a new dog food on Chewy.com, but when you leave that site and start browsing other sites, you begin to see promotional ads for dog food on the sidebars. When a company starts using this type of targeted internet advertising, better results are seen, and more people are reached for your specific product or service. This type of marketing and advertising is driven by artificial intelligence.

Another factor to consider is cost. It turns out that advertising on social media platforms can be dirt cheap. You can start advertising on Facebook in order to do testing of advertisements for as little as five dollars a day. No, of course, you're not going to advertise at that lower level if you're looking to scale your business to reach large numbers of customers. Still, you would be surprised at how effective this is for testing ads to determine which ones are going to work or not.

So, we've already identified Facebook as a perfect marketing platform. Facebook lets you finely target your advertisements to the specific demographics most interested in your product, service, or business. Another advantage of Facebook marketing is that you are going to be able to collect a lot of data in the process. The data that you collect with Facebook marketing is going to be something that you can use later to make your advertising efforts even more focused. Don't forget the case study that we mentioned earlier with the look-alike audiences. The conversion rates that you get with Facebook lookalike audiences (created using data together with machine intelligence) are sky-high in comparison with standard conversion rates. It would be interesting to compare the conversion rates from Facebook look-alike audiences to those that are obtained by throwing an ad up on a website like CNN.

Let's consider another social media website that we discussed in the last chapter: YouTube. YouTube is one of the busiest sites on the planet. Although it's not immediately obvious, you can use YouTube for rock bottom priced advertising in order to target people who have specific interests. All you have to do to advertise on YouTube is to create a simple

video and then use that as the basis of an advertisement. Then you can state specifically where are you want that ad to show. You can even have them show it preceding the viewing of specific videos that you have identified. So, if you got a product to solve a certain need, you could have it advertised on videos that discussed a related issue. As an example, let's say that you sold a portable personal EKG machine. So, you can make a video demonstrating the machine to use as an advertisement. Next, have that ad placed just before videos about heart attacks and symptoms of heart attacks. As you can surmise, the people who are coming to view the video are already worried about having heart attacks. When they click to watch on the health video, your ad shows up, demonstrating the personal EKG machine for them to use. You can imagine how well that kind of targeted advertisement would convert!

YouTube is also a great site to utilize organic for free marketing. This does take a lot of time and effort, however. But if you're careful about the keywords you use in the video topics that you choose, making professionally looking videos over time, you will be able to build a following on YouTube and market to them. The key to good marketing videos on YouTube is to make the videos as if they were not marketing videos. You don't want to be explicitly marketing to people on YouTube. What you want to do on YouTube is to show people how to solve problems that are related to your business. This is like the approach when using a Facebook page. On the Facebook page, the apparent purpose of the page is to develop relationships with customers, not advertising to them - at least not very often. Building relationships with customers is something that's very effective because it creates a level of trust. Once people trust you, they are going to be easy to sell to and turn from prospects into customers of your business.

## **Customer Targeting**

The amount of data that is available on social media is enormous. However, within the data, it is possible for companies to extract all kinds of information. You can learn things like education level, page likes, marital status, number of children, and what websites they visit.

This can have a lot of power when it comes to marketing. One of the key characteristics the social media provides is the ability of businesses to gain insight into customer demographics. This can be turned into actionable decisions by the business. This can include more effective targeted

advertising, and the business can engage in active involvement with a potential customer base within the social media platform.

This offers an opportunity for the most efficient marketing and sales possible. First, a company can use Facebook or other social media companies in order to determine the demographics and interests of its customer base.

## **Search Engine Marketing**

Another way that companies can use their online presence - together with large amounts of data - is through search engine marketing. You can use data that has been collected by companies like Microsoft and Google in order to laser target potential customers as they search on the Internet for topics related to your business. This type of marketing is powerful and inexpensive. Some of the data that you can utilize includes keywords from searches. Search engines will allow you to download data that includes keywords together with the number of searches that are done on each keyword. This helps determine which keywords are best to use in your own advertising.

## **Establishing an Internet Presence**

Being everywhere is a critical component of your internet marketing efforts. This will involve new media and older media. By this, we mean that you should be on relevant social media platforms, providing real, relevant, and engaging content that helps you form relationships with customers. Second, you should have a blog. Many businesses miss the importance of a blog, not to mention the opportunity that exists for the collection of more data. Your blog should also engage prospects with timely, interesting, and relevant content that will help to expand your presence online. At the same time, you can collect data from readers while speaking to them directly. It can also be integrated with your Facebook page in order to create a feedback loop that can help power your online presence and marketing efforts.

## **Using Business Intelligence to Analyze and Improve**

After you begin a large-scale Internet marketing approach, you will be able to turn your advertising and marketing data into inputs to business intelligence. This will allow you to convert raw data (about who responds to your marketing efforts) into reports and visualizations which can help management make data-driven decisions on future marketing efforts.

Another way to handle the data from Internet marketing is by feeding that data into machine learning systems. This is going to be useful to detect patterns, trends, and relationships in the data that can be used to gain actionable insights for the next round of your internet marketing efforts.

The key to a winning strategy with internet marketing is not to do more of the same. Instead, you should be continually collecting data from internet marketing efforts and feeding it to business intelligence, creating updated information and reports that can easily be absorbed and analyzed. The amount of data that is online is massive, and even in your own marketing efforts, you are going to be collecting colossal amounts of data that can seem to overwhelm. However, by using the right business intelligence process together with big data and machine learning, you are going to be able to turn that data into usable, actionable information that will take your marketing efforts to another level. This is going to be a continual feedback loop that will constantly be taking in enormous amounts of data and processing it in ways that can help your business refine its marketing efforts.

## **Identify Weak Spots**

One of the most important uses of business intelligence with data collected from internet marketing is going to be identifying weak spots in your marketing plan. By processing, analyzing, and presenting data in a readable form, you will be able to find out what marketing efforts are not working and why. You will be able to update, refine, and improve your efforts and cycle through the process in order to get better results. When marketing efforts fail, you can kill the ones that don't work and put more effort into those that do. Each time you go through the cycle, you will achieve stronger and better internet marketing tactics that work to bring results.

## **Is internet marketing for everyone?**

For some types of businesses, one question that arises is: should everyone be marketing online? The answer to this question is an absolute "YES!" The reason is simple: the data that can be collected is going to provide your business with new insights and actionable information. Even small "mom and pop" businesses would benefit from collecting and analyzing online data. Obviously, they are not going to have the infrastructure to do that themselves, but they can utilize services from companies like Facebook, gaining insights into their customer base and how they can better reach them. The companies that are constantly upgrading and updating are going to be the



ones that will be able to achieve a competitive edge helping them achieve and maintain dominance in the marketplace.

## Conclusion

Thank you for taking the time to read this book about business intelligence. If you did not know what business intelligence was before you started this book, it probably sounded like nothing more than another buzzword that has permeated corporate America. However, contained within the hype, business intelligence describes the kind of useful and important processes that businesses must adopt in order to stay competitive in this new technological and data-driven environment.

Business intelligence has been with us for some time. As soon as data was able to be gathered electronically and organized into the form of relational databases, businesses began developing procedures that would come to be later recognized as business intelligence. The advent of big data in machine learning is going to drive a revolution in business intelligence. Big data is completely changing the way that business intelligence is going to be used in the future. Big data represents a massive shift away from structured data in the form of relational databases and spreadsheets, and toward unstructured data that is being collected in massive amounts across all aspects of the business enterprise. Furthermore, big data can be used both for customer behavior and for the internal behavior of the employees of the company. Huge treasure troves of data have been collected on employee behavior, reflecting everything from emails, word processing documents, meeting notes, images, videos, and slideshow presentations. There are also huge amounts of data related to the use or misuse of resources within the company. One example of this involved Southwest Airlines trying to discover the reasons for – and costs of - leaving an idling aircraft on the tarmac. By using machine intelligence, the company was able to identify the problem and develop working solutions that were cost-effective, saving the company hundreds of millions of dollars in fuel costs. This example serves as a classic application of big data and machine intelligence for machine learning to business intelligence. Before we mentioned fuel savings, identification of the problem, and large cash savings by the business, you might not have identified machine learning as part of business intelligence. However, when we bring all these things together, it makes it clear that machine learning and big data have a direct application to the types of problems that are routinely solved using business intelligence.

Data has not only been available from internal sources collected by the company itself. In fact, data is being described as the new currency of

business in the 21<sup>st</sup> century. As a result, many companies (including large ones like Facebook and Google) are using data itself in order to derive value for the company. This data can be shared with other corporations to drive the improvement of their own business intelligence.

As big data continues to be collected at a record pace and machine learning continues to improve and become more and more incorporated in business, it is incumbent upon corporations of all sizes to incorporate machine learning and big data into their strategy for business intelligence. Those companies able to do this successfully are the companies that are going to have a large competitive advantage in the marketplace. Those companies that are unable to do this (or that neglect to do it) are going to be the companies that are left behind. Thought processes must change when it comes to updating business intelligence. No longer can a company simply rely on the structured data that they have compiled into relational databases. As we mentioned earlier in the book, it is estimated that 90% of data that is now collected is in the form of unstructured data. Many companies are lost, not even sure how they can use this data or what they can do with it in a cost-effective manner.

As we go forward into the era of big data, cybersecurity is going to be increasingly important. Not only is it going to be important in a general sense, but cybersecurity is also going to play a larger and larger role in protecting business intelligence. Malicious actors are becoming much more emboldened as a business becomes more reliant on collecting and maintaining data. Therefore, although cybersecurity is not directly considered within business intelligence, it is going to play a supportive role that becomes more crucial as time goes on.

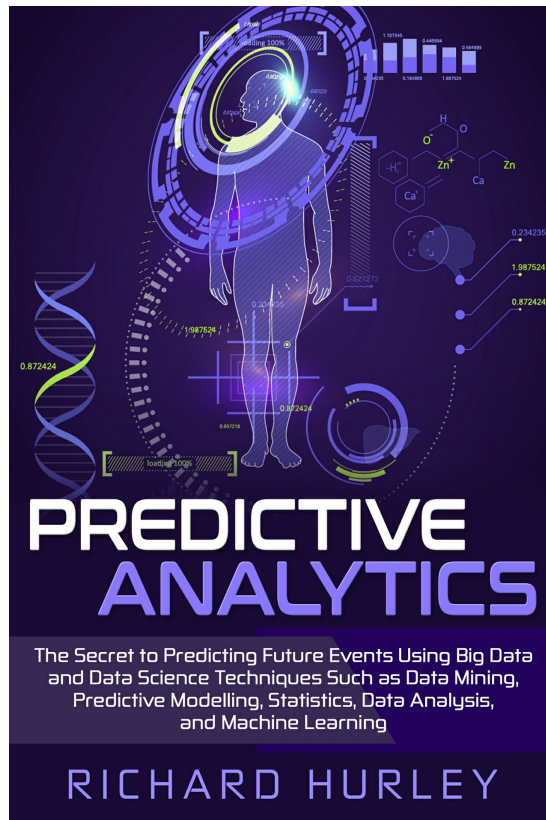
Many businesses have also been slow to adapt to the social media landscape. Over the past decade, social media has evolved from being a fringe activity of young people to becoming a centralized and crucially important component of the entire media landscape. Many issues are clouding the social media arena right now, including issues of privacy - which could have a large impact on the ways that businesses are able to share data amongst themselves. But one thing that we can be sure of: social media is here to stay. In one form or another, it will play a central role in communication and marketing in the future.

At this point, I would like to thank you for taking the time to read this book. I hope that you have found it informative and educational. If you have enjoyed the book, we certainly would appreciate it if you would leave us a

review on Amazon. I would also like to encourage you to continue your education on the topics of big data, data science, machine intelligence, and business intelligence. Good luck in it all!

## Part 2: Predictive Analytics

*The Secret to Predicting Future Events Using Big Data and Data Science Techniques Such as Data Mining, Predictive Modelling, Statistics, Data Analysis, and Machine Learning*



# Introduction

The following chapters will discuss everything that you need to know when it comes to what predictive analysis does and why we all should learn how to use this process when it comes to our raw data. There is an enormous amount of raw data coming to us from all angles, data that businesses are taking in all the time. But the important part here is not how much data they can take in, but how they are able to use this data to learn something new, to find new trends, and to make predictions on how they should behave and act in the future. This guidebook will help you understand one of the processes that you can use - the predictive analysis - that will help when it is time to make some of these predictions.

The beginning of this guidebook explores the basics that come with predictive analysis. We will look at how this works, and why it is so beneficial to use this process. We can then take a look at some of the different methods that you can use during your predictive analysis in order to make these big predictions, including data mining, big data, predictive modeling, statistics, and data analysis.

Next, we'll move on to a few of the other topics that are going to come into play when it is time to work with predictive analysis. We will add in a bit of an introduction to machine learning and how this can help with making predictions, and some of the options that you should follow when you want to avoid the prediction traps, the biases, and more that can interfere with the predictions that you want to make with these processes.

Towards the end of this guidebook, we are going to spend some time looking at the best steps that you can take to make sure that your predictive analysis is as successful as possible. There are a few ways that we can do this, but we will start with some of the top reasons to use predictive analysis in your business and why it is so beneficial, and some of the top steps to help you create one of your own predictive analysis models to help take that raw data and make it as useful as possible in your life.

There are a lot of different methods that you can implement to go through all of the raw data that your company regularly retains. Rather than trying to guess the kind of decisions you should make without any proof or data behind these decisions, why not use predictive analysis to help avoid this risk and improve your business decisions? When you are ready to learn more about predictive analytics and how it can work for your needs, this guidebook will show you how to get started.

# Chapter 1: The Basics of Predictive Analysis

The first topic that we need to consider is predictive analysis. This topic will be the common thread that we explore throughout this guidebook. As such, having a good understanding of what this all means and how we can use this to our advantage is going to be so important to how well we can do with some of the other topics in this guidebook as well.

To start with, predictive analysis is helpful when we use data, machine learning techniques, and other statistical algorithms to help us identify the likelihood of an outcome in the future, usually based on some of the data that we have from the past. The goal of using this is to go beyond just knowing what has happened so that we can then come up with a good assessment of what we think will happen in the future.

Of course, this is not always going to be accurate all of the time. Sometimes, the future doesn't behave in the way that we would expect, or some big changes in the economy or the market have delivered results that no one is expecting. But the idea here is that we want to use this kind of analysis because often, the way the company has done in the past is like how they will do in the future, as long as all things remain constant.

This could be used when you are trying to figure out how to schedule employees based on sales, for example. If it is a cold winter, and you had lower sales the last few colder winters, then it is likely that you will not schedule as many people to handle all the shifts because there won't be as much traffic. If you make some big sales around Christmas time, then you would want to make sure you schedule more shifts at the right times to handle this higher business level.

Though the work of predictive analysis is something that we have seen around for decades, it is a technology that is starting to seem more valuable recently. There are a lot of organizations throughout all the different industries which are seeing the value of this kind of analysis, and they are turning to it to see how they can benefit. The point of doing this is because it provides the company with a big advantage over the competition, and it can even help their bottom line. But this brings up the question, why is the idea of predictive analysis becoming so popular now, even though it is something that has been around for quite some time?

There are actually a couple of different reasons for this. Some of these reasons include:

1. There is a growing amount of data types that companies are using and gathering now. In fact, many companies have more data now than they have ever had in the past, and they need to know how to use it. There is also a growing interest in using this data in an efficient way to produce insights that are valuable and can help the business to perform better.
2. Computers that are faster and less expensive.
3. Software that is easier than ever before to use on a variety of tasks.
4. Economic conditions that are tougher, which means that each company - no matter what kind of industry they are in - needs to be able to differentiate themselves from the competition.

With software that is interactive and easier to use than ever before, and with the fact that this is becoming more prevalent all the time, predictive analytics is no longer just the domain of statisticians and mathematicians, as it was in the past. Line-of-business experts and business analysts are now opening and using a lot of these same technologies for their companies as well.

We can also take a look at why this predictive analytics is going to be so important. Organizations are going to turn to this analytics in order to help them solve some difficult problems, and uncover some new opportunities. This means that there is already a lot of opportunities out there for this kind of process already.

With this in mind, some of the common uses that we are going to see with predictive analysis will include the ability to detect fraud. Because predictive analysis is able to combine together multiple methods of analytics, it is able to improve the detection of patterns and prevent a lot of criminal behavior. As the world is on the lookout for more types and more advancements in the cyberattacks that come, being able to have the right kinds of analytics in place can make sure that all the abnormalities that are present can be caught ahead of time and can keep personal and financial information as safe as possible.

Another benefit is going to come in the form of optimizing the marketing campaigns. Predictive analytics will help companies and marketers determine the purchases or the responses of the customer, as well as help promote some of the cross-selling opportunities as soon as they come up. Predictive models are going to be able to help with this because they are good for assisting a



business in attracting, retaining, and growing some of their customers who are the most profitable.

You can also use these predictive models to help with improving operations. Many companies are going to use these predictive models to forecast inventory and manage the resources that they have. For example, this is something that you will see with airline tickets because they will use these models to ensure they set the ticket prices at the right point, based on the demand at that time of year. Hotels can do this by predicting how many guests are going to show up on any given night, maximizing occupancy and increasing their revenue at the same time. When predictive analytics is used properly, it is going to enable the organization to function in a manner that is more efficient, increasing their bottom line.

We can see how predictive analytics can come into play and reduce the amount of risk that the company is going to assume. For example, credit scores are often used because they can do an assessment of how likely it is a buyer will default on their purchases, and it is one of the best-known examples of how this predictive analysis is going to be able to work.

The score that you get on your credit is actually a number that has been generated through the use of predictive analysis, based on some of your past habits with making purchases, asking for and either receiving or being denied for a loan when you apply, whether you have made payments on time in the past and more. Other risk-related uses of this kind of model are going to include things like the claims and the collections that you try to do with your insurance company over time.

With this in mind, we also need to take a look at who is using this kind of model in order to gain a competitive advantage over others in the same industry. Pretty much any industry will be able to use this predictive analysis to help them to increase their revenue, optimize how they run the company, and to reduce the risks that they face. But there are a few industries already jumping on board and seeing a lot of benefits in the process. These industries are going to include:

**Banking and Financial Services.** The financial industry is often going to handle a huge amount of money and data on a regular basis. In between people putting money into and taking money out of their accounts, the various loans they offer, and more, it is no wonder that this industry is going to use some of the predictive analytic models to help them out. They can use these models to help out with the detection and the reduction of fraud, to help

measure the credit risk of various customers, to maximize the cross-sell or the upsell opportunities, and to make sure that they will retain valuable customers along the way.

One example of this is done with the Commonwealth Bank. They have been able to utilize the predictive analysis models in a way that they can figure out the likelihood of fraud pretty quickly. This company can predict how likely it is that fraud is occurring for any given transaction before that transaction has been authorized - *within 40 milliseconds of the initiation* of a particular transaction.

Think of how this can change the financial world. It builds up a lot of trust with the bank because it is more likely they will catch someone who is trying to make fraudulent purchases long before they can get away with anything. It helps to put the mind of the customer at rest because they know their money is safe, and often the check for fraud is done without anyone even noticing because it is that quick. The bank is also going to benefit. They can save billions of dollars in the process if they can stop fraudulent charges before they even happen, saving time doing investigations into the transactions, and the potential for losing the money because the person who did it was never caught will go down to a minimum as well.

**Next on the list: Oil, Gas, and Utilities.** Whether these companies are predicting when equipment will fail so they can get things fixed or replaced quickly without a bit interruption, or they are predicting the future needs of their resources, mitigating safety and reliability risks, or trying to improve their performance overall, the energy industry has really taken a liking to predictive analytics and using it on a daily basis.

A good example of how this has been used is with the Salt River Project. This is the second largest of the public power utilities in America, and one of the largest suppliers of water in Arizona. Analysis of machine sensor data is able to help this company predict when their turbines, the ones that give power, need some maintenance along the way.

**Retail companies** can benefit from this kind of analysis, as well. At one point, a study that is now infamous shows that men who go to the store to buy diapers also will purchase some kind of beer at the same time. And since this time, retailers have been using predictive analytics to help them plan out their merchandise and optimize the prices that they sell things to. This allows them to see how effective their promotional events can be, and will make it easier to determine which offers are the best to start.

While many retailers do this on a regular basis, we are going to take a look at Staples for a moment. Staples was able to use predictive analysis to gain customer insight. They analyzed behavior, which gave them a complete picture of their customers. When they were able to put this into practice, it resulted in a return on investment of 137 percent.

We can also see **governments and the public sector** using this kind of technology. Governments have already been one of the key players when it comes to the advancement of computer technologies. For example, the US Census Bureau has been analyzing data to help understand trends in the population for a long time. To add to this, governments now use predictive analytics like many other industries so that they can improve their performance and their service, so they can detect and then prevent fraud, and help them better understand the behavior they see with the consumer. They can often add in cybersecurity to their use of predictive analysis.

**The health insurance industry** uses predictive analysis as well. This can be a risky business because you want to provide the best service to customers, but you must balance out the costs that it takes to help patients with the amount that you are receiving each month. And predictive analysis can really help to make all this fit together.

First, predictive analysis is going to help when it is time to detect fraudulent claims. This industry is taking it a bit further, though, by taking steps to identify patients who are at most risk of chronic disease and then find out what kinds of interventions are going to be the best to help them out. This can help keep the patient healthier while saving the insurance company a lot of money in the long run.

For example, Scripts, who is a large pharmacy benefits company, has used predictive analysis. They use these analytics in order to identify those who are not adhering to their prescribed treatments, which has helped them to save somewhere between \$1500 and \$9000 per patient. This can help the insurance company lower costs while still giving their customers the healthcare and medicines that they need.

We can also see some of this predictive analysis show up in the world of manufacturing as well. For these companies, it is important for those running it to identify factors that could lead to reduced quality and even to failures in production. They also need to focus on optimizing the parts, the service resources, and the distribution.

One example of the world of manufacturing using this analysis is with the

company Lenovo. This is just one of the manufacturers recently who has started to use these predictive analytics in order to understand more about their warranty claims. This was a big initiative in the company that led to an up to 15 percent reduction in warranty costs, saving a lot of time and money overall for the company.

With this in mind, we need to look at some of the information about how the predictive model is going to work. These models can take known results and use them to develop - or train - a model that can be used to help us predict values for different or new data that shows up. Modeling is going to provide us results in the form of a prediction, which will just be the probability of the target variable, based on the estimated amount of significance from a set of input variables.

This is actually going to be a bit different from what we will see with descriptive models. With the descriptive models, we use them to help us understand what happened. And then there are diagnostic models that are a bit different as well because they will help us to understand some of the key relationships that go on in the business, and then we can use this to determine why something has happened. There are entire books that have been devoted to analytical methods and various techniques, as well. But we will focus on just some of the basics that come with this to help us see how it all works.

The first thing to explore here is the two different types of predictive models that we can work with. The classification models are going to be able to predict class membership. For example, you may want to do a model on whether any of your employees are likely to leave, whether a customer is going to respond to a solicitation, and whether the customer is going to be a bad or good credit risk for you.

Usually, the model is going to give us results that come in the form of 0 (zero) or 1 (one) so that it stays binary and is easier to work with overtime. 1 is going to be the event that you are targeting, and 0 means that the event you are targeting is not going to happen.

Then we can also work with the regression models. These are the models that are able to predict a number for us. This means that they can help us see how much revenue a customer is able to generate over the next year, or the number of months before we think a component is going to fail on one of our machines, so we can make sure to order it and get that part fixed in a timely manner without stopping production.

With this in mind, there are also a few different techniques that you are able

to work with when you use predictive modeling. The three most common options are going to include *neural networks*, *regression*, and *decision trees*. To start with is the decision tree. These are going to be models of classification that can partition your data into various subsets based on the categories that you use for the variables of input.

This is going to make it easier to understand the path someone took to make their decisions. A decision tree is given this name because it looks like a tree with each branch representing a choice between a number of alternatives, and each leaf representing a classification or a decision. This model can be useful because it will look at the data and then tries to find the one variable that can split that data into groups that are the most different.

*Decision trees* can be popular because they are easy for us to interpret and understand. They also will help you handle any missing values that show up, and can be useful for preliminary variable selection. So, any time you are working with a set of data and you have a value that is missing, or you would like to work with an answer that is quick and easy to interpret, you can start making a decision tree.

You can also work with *regression* when creating a predictive model. Both logistic and linear work here. This is actually one of the most popular out of the methods in statistics because it is going to help estimate the relationships that are there between variables. It is intended to use with continuous data that we can assume is following a normal distribution, and it can help find key patterns in the larger sets of data. And sometimes, a programmer will use it to determine how much specific factors, including the price, will influence the movement we see with an asset.

When you use the regression analysis, you want to be able to predict a number, and then the response of this is going to be the Y variable. With the linear regression option, one of the independent variables will be used to help explain and predict the outcome you will get with Y. Multiple regression can come in as well and we will use two or more independent variables to predict the outcome.

With logistic regression, though, the unknown variable of a discrete variable is going to be predicted based on known values of the other variables. The response variable is going to be categorical, meaning that it is only able to take on a limited number of values. When we do the binary logistic regression, a response variable only has two values, including 0 or 1. In multiple logistic regression, a response variable can sometimes come with

several levels, including 1, 2, and 3, or low, medium, and high

And the third type of model that you can work with is the *neural networks*. These are going to be sophisticated techniques that are capable of modeling relationships that are really complex. These are going to be popular in this kind of field because they are so flexible and so powerful in the process. The power with them is going to come in because they have the ability to handle any of the relationships that show up in the data that are nonlinear, which is going to show up more and more as companies work to gather more data from their customers, and from other sources.

These neural networks are going to be there to help us to confirm the findings that we may see with some of the other techniques, including the decision tree and the regression. They are going to work because they are based on pattern recognition and some of the artificial intelligence processes to model the parameters that you would like graphically.

They are going to work well, even when you do not have a known mathematical formula that will relate to the input or the output, and when the prediction is going to be the most important thing in the mix – even more important than the explanation. You can also work with the neural network when you have a good amount of training data to use in teaching the algorithm or the method the right way to behave.

Knowing a bit of artificial intelligence with this can make things easier because a lot of the methods that work with this idea, including machine learning, can make the neural network stronger. In fact, the artificial neural networks were originally developed by researchers who were trying to mimic the neurophysiology of the human brain.

These are the most common types of techniques that you can work with when it comes to doing predictive analysis, but there are certainly not the only ones that you are able to work with as you do your own predictive modeling. Learning about the different methods that you can use, and some of the times when you would pull each one out can make a big difference in the amount of success you are going to have with these predictions. We will take some time to discuss a few more of these methods, as well as discussing in more detail the ones above as we go through this guidebook.

## **Chapter 2: How to Predict Future Events with the Help of Big Data**

Now that we have some idea of what the predictive analysis is all about, and why it is so important for helping out a business, it is time for us to look at big data. We did talk before about how data is really important in this kind of analysis. Without the data, we have nothing to base our predictions on in the first place. Knowing what is inside the data, and then using it to make decisions and predictions are what predictive analytics is all about. With that in mind, we need to take some time to understand Big Data and why it has become such an important part of many companies.

Big data is not a concept that is hard to understand or going to be complex. It is simply the term that we use in order to describe a large volume of data. Sometimes this data is going to be structured, and other times it is going to be unstructured. This is the data that will come into a business on a day to day business. Some companies have a lot of data, and some are not going to have as much, but the interesting thing here is that the amount of data is not really that important.

What is important, and what we will need to focus on here when we do predictive analysis, is what the company is going to do with the data, big or small, that really matters. Yes, Big Data usually refers to a large amount of data that the company has. But we really need to focus on how that data is handled and used by the company to see what will happen with them.

A company can have all of the data in the world. But if they choose not to use it, just store it away for later, or if they don't understand the data, then it doesn't do them any good at all. They may as well have no data on them at all. This is where the predictive analysis is going to come into play, though. It helps the company to take all of the data they are holding onto, and analyze it for various insights and trends that may be hidden in there, leading to the best decisions and strategic moves for the business.

The term "Bog Data" is a relatively new concept, and it wasn't heard in the world of business until the last few years; we can all agree that companies have been gathering, as well as storing, large amounts of information to analyze at some point for a long time. This is how they come up with their current advertisements, their promotions, their new products, and more - all in order to entice their customers to make a purchase.

It was in the early 2000s when this concept started to gain some momentum.

This happened when Doug Laney, one of the analysts in the industry, articulated a now mainstream definition of what big data is all about. Remember, it is not necessarily just about the amount of data that you are holding onto; it is about a few other things as well. The best way to remember what Big Data is about is to remember the “three Vs” which include the following:

1. **Volume.** Organizations are able to collect data from a large variety of sources, which can include information they get from their machine to machine data, from their various social media accounts, and from business transactions. In the past, they may have limited the amount of data they take in because of the storage, but with some of the newer options available, this is a burden that is a bit easier to handle.
2. **Velocity.** Data is able to stream into a company at speeds that are unprecedented and hard to keep up with. Even though it is a ton of data that is coming in quickly, it still has to be dealt with in a manner that is timely if the company wants to still gain some insights out of it. Smart metering, sensors, and tags are helping many companies to deal with this information in a timely manner, at least as much as possible, so that companies can keep up and gain the right predictions out of that information.
3. **Variety.** We have to remember during this process that variety is going to be important with the data that we use. Data is going to come to our direction from lots of formats. This can include text documents that are not structured, numeric data that is found in a traditional database, and structured data as well. We can also see things like financial transactions, stock ticker data, audio, email, and video, to name a few.

It is the responsibility of the business to be able to handle all these varieties of data and learn the best way to get them to conform with one another to give you results.

We can also take this idea a little bit further. In addition to the “three Vs” that we talked about before, it is also possible for us to use a few other ideas to help understand Big Data and what it is all going to be about. The other two parts of Big Data that are important will include:

1. **Variability.** In addition to the increasing varieties and



velocities that happen with the data, the data flow can be inconsistent throughout the month or year. There are going to be times with higher flow, and times with lower flow. Daily, event-triggered, and seasonal peak data loads can be hard for us to manage sometimes. And this becomes an even bigger problem when we deal with data that is unstructured. Having a plan in place and learning when these peaks can help your business to thrive.

**2. Complexity:** Today's data is going to come from multiple sources, which is going to add in another layer of challenge for us to link, cleanse, match, and transform data across the systems. However, it is important for a company to be able to connect and then correlate relationships, multiple linkages in data, and hierarchies in the data – or that data is going to start spiraling out of control!

This brings us to the next point that we need to consider. We need to take a look at why this Big Data is so important, and why we are taking some time now to discuss it and learn more about it. The importance of this big data is not going to revolve just around how much data you are holding onto, but what you plan to do with that data. You can basically take that data from any source that you would like, and a few things are going to happen.

When you analyze the Big Data, you will find that it can help you get answers that enable a reduction in costs for the company, a reduction in the amount of time that it takes to get things done, the ability to develop some new products, and optimized offerings, and it can increase the number of smart decisions that you can make.

When you can combine this Big Data together with high powered analytics or something like predictive analysis, you are going to be able to accomplish a wide range of business-related tasks. Some of these tasks include:

1. Helping you to determine some of the root causes of failures, defects, and issues in real-time, making it easier to fix the issue and get things back on track.
2. Monitoring the customers you have and then generate coupons at the point of sale. These coupons are going to be based on the buying habits of the customer.

3. Looking at the portfolio that you have and then recalculate the amount of risk that is found there in just a few minutes rather than in hours.

4. It is great at detecting behavior that is more fraudulent before it is going to have any kind of effect on your organization.

Big data is going to have a big effect on all organizations, no matter which industry you are looking at. In fact, all of the industries that we talked about before when it comes to the predictive analysis are also going to benefit from the use of Big Data. Plus, there are many others who are on board and seeing the benefits as well. Let's take a moment to look at the industries that use Big data on a regular basis now, and how they are able to benefit from all of this new information coming in.

First on the list is banking. Just think about the number of transactions that banks and other financial institutions have to deal with on a regular basis. They have to keep track of debit and credit card purchases, people depositing money into their accounts, loan information, and more. And they have to make sure that no fraudulent activity is going to happen in the meantime. This can make it hard to keep up and to make sure that the money is safe and sound throughout the day.

Big Data is going to come in because it helps with all of this. This brings in a lot of big insights on the customers that the business has, while also requiring that the financial institutions always stay a step ahead of the game when it comes to advanced analytics. The bank or financial institution will be able to use the Big Data to help them to pick out the right individuals to give loans to, to watch for activity that may be criminal in the purchases, and more, saving them a lot of money.

Education can use this Big Data as well. Educators, when they are armed with insights that are driven by all of that data, are able to make a huge impact on the school system, the curriculum that they work with, and their own students. This Big Data is also going to come into play when they can analyze it and then identify students who may be more at risk than others.

With this information in hand, the teacher and others at the school are able to find the students who need the most help, watch to make sure that they are making good progress, and even come up with a better system of support and evaluation for principals and teachers to ensure that everyone is on the same page the whole time.

Another industry that is going to benefit from all of this Big Data is health

care. Think about how much information goes through a hospital or a clinic. Many times a patient is going to have at least a few records just for going in for their regular checkups. Then if they ever have surgery, give birth, have to get a scan done, get referred out to a specialist, change insurance, or anything else, this is more information that the hospital or the clinic is going to need to take care of as well.

Between prescription information, plans on treatment, and patient records, everything has to be done in a quick and accurate manner in this kind of industry. There also has to be enough transparency that goes on to satisfy all of those stringent regulations in the industry. When this data is handled in the right manner, the providers can do their job better, and can even uncover some of the insights that are hidden in all of that information, improving the care for the patient.

The world of retail is another place where we can use this big data as well. Customer relationship is going to be so important when it comes to the world of retail, and the best way to make sure that the company can improve that relationship is for them to manage big data and find the insights that are hidden inside.

Retailers have to know the best way that they can market to their customers, the most effective ways that they are able to handle the transactions, and even the most strategic methods they can use when it is time to bring back any business that has lapsed over time. And at the very heart of all of these things, we are going to find Big Data.

Before we move on here, we need to take a look at how big data is going to work. Before we can really see all of the ways that your business is able to benefit from big data at work, we first need to have a better idea of where this data is coming from in the first place. The sources of our big data can come from a lot of different sources, based on what information you are trying to collect. However, the sources that most companies and industries are going to use when bringing in the big data can fall under one of three categories, and these categories to include:

1. **Streaming data:** When we look at this category, it is going to include any of the data that is able to reach to your IT system from a web of connected devices. You are able to take a look at and analyze this data as it arrives, and it will make decisions on what data you would like to keep, what you don't want to keep, and what may need you to analyze it a bit further than before.

**2. Data from social media:** The data on social interactions is a part that is starting to look even more attractive to many companies, especially when it comes to support, sales, and marketing functions. This kind of data is going to come in the semi-structured or unstructured forms, so it is going to pose a new challenge when you want to use the data or analyze it at all, but it is still useful when you know how to make it show you some good insights.

**3. Sources that are publicly available:** There is a massive amount of data that is available to companies through open data sources - looking for the ones that are available in your industry (and with topics that you are the most interested in) can help you to really see some good information.

After you have had a chance to identify all of the different sources of data that you would like to use, it is now time for you to consider the decisions that you want to make after you harness in all of this information. These do not have to be complicated, but it is still important to have a plan for any and all of the big data that you plan to gather and use for your needs. Some of the decisions that you have to consider and make when it comes to the Big Data you are using will include:

1. *How you would like to store and manage the data.* A few years ago, the storage of all this data was a big issue that companies did not know how to handle. The good news is that there are a lot of low-cost options to hold onto that data now. If this is what you would like to do with the data, then you need to take some time to research and figure out which storage is right for you.

2. *How much of the data you would like to analyze.* This is going to be different for each company based on what they are hoping to get out of the information and what they plan to do as their strategy. Some organizations decide to not exclude any data out of their analysis. This is something that you can do thanks to the high-performance technologies that are out there, such as in-memory analytics or grid computing. But for some companies, it is not necessary to look through all of the data, or it may be too expensive to do this. So, they decide to look at the information and decide which part is relevant, and which part is not before starting

the analysis.

3. *How to use the insights that you uncover.* The more knowledge that you have present, the more confident you will be when you make these important business decisions. It is really smart for you to have a strategy in place after you have been able to gather an abundance of information to use.

The final thing that we need to do when it is time to take our Big Data and put it to work for our business is to do some research. This research needs to be done on the various technologies that are available and will help you to make the most out of your big data and can be helpful for your predictive analysis and your big data analytics. Some of the things that you need to consider when you go through this part of the process include:

1. Most of the storage that you can use is cheap, but make sure there is enough room for what you would like to do.
2. You want to work with processors that are faster so it can keep up with this information.
3. You want to go with technologies that are affordable and fit in your budget, open-sourced, and are going to be distributed with the help of big data platforms.
4. You want to consider something like clustering, virtualization, and parallel processing based on how much data you have and what your end goal with all that data is.
5. Many companies also like to work with cloud computing or some other kind of resource that is flexible and allows for arrangements of allocation. It is going to depend on what your end goals are with the data when it comes to which one you would like to use.

The neat thing that comes with big data is that you are able to use it to make the predictions that you want about the future. There is so much information that you are gathering with this Big Data, and the whole point of it is to use this information to learn something about your customers, about your business, and even about your competition and the industry that you are in. If you can use this data in the proper manner, you will find that it is easier to make predictions about what is going to happen in the future.

In fact, many companies are already using this Big Data in order to help them

learn insights about their business, and then to make predictions on how things are going to happen in the future. Think of it this way; would you rather just make random decisions and shoot in the dark, hoping that you are going to make the right predictions, or would you rather have a large source of data behind you that points in you in the right direction for the decisions and predictions you make? If you are going with the second choice, then this is exactly where Big Data is going to come in and make a difference. As long as it is used properly, and you have done the right predictive analysis on it, you will find that making predictions with the help of Big Data can be easy.

## **Chapter 3: Can I Predict Future Events with Data Mining?**

Another process that we need to explore a little bit while we are here is the idea of data mining. This is a very important piece of the puzzle when we talk about predictive analysis, and it is going to be so helpful when it is time to make some predictions about whether an event is going to happen with your customers, with your industry, or with your business in the future. Gaining a better understanding of what data mining is all about, and why it is so important can be the key to ensuring you complete your predictive analysis in the proper manner in this process.

To start, we need to take a closer look at what data mining is all about and why it is so important. Data mining is going to be the process of finding anomalies, correlations, and correlations within a large set of data in order to predict the outcomes that are the most likely to happen. There are many techniques that will come into play here, and when you use them properly, it is possible to increase your profits, cut out on costs, improve the relationship that you have with your customers, lower the amount of risk that you are taking on, and so much more as well.

The process of going through all of the data that you have in order to discover some of the hidden insights - then using these insights in order to predict some of the future trends - has been a process that companies have done for many years. In fact, it has been researched quite a bit to see whether this method is effective or not, and it is going to include three intertwined disciplines of science, including machine learning, artificial intelligence, and statistics.

What do all of these mean? Statistics is going to be the numeric study that we can do in order to see what relationship is there in the data. Then we have artificial intelligence, which is when we see what is like human intelligence, displayed by the machines or the software that we use in data mining. And then, machine learning, which includes all the models and algorithms that we need in order to help look through the data and make predictions. All of these come together to help us through the process of data mining and making predictions based on the information we collect.

Over the last few years, we have also seen that our systems and machines have had some big advances in processing power and speed, and this has allowed us to move beyond some of the time-consuming, tedious, and manual

processes that we did in the past. And in its place, we are now onto a data analysis that is automated, quick, and easy, which makes things better for everyone who is involved.

An important thing to note: the more complex the set of data that you can collect, the more potential that you must uncover insights that are relevant and helpful. Yes, it may take longer to get through all of this data, but it means that you are going to find some more insights that can really propel your business forward, and give you a competitive edge over others in the same industry. With more data, it is easier to know what promotions to do, how to reach your customers, the demographics of your customers, and more. So, why is this process of data mining so important? If you have been in business for any length of time, it is likely you have already seen the numbers, and you know the fact that the volume of data that is produced is doubling every two years. Unstructured data - on its own - can make up at least 90 percent of what we see in the digital universe. But we always have to remember that more information is not going to be the same thing as more knowledge.

When you work with data mining, it is going to allow you to do a few important things. These are going to include:

1. Data mining can make it easier to sift through all the noise, both the chaotic and the repetitive, that is found in any set of data that you want to work with.
2. Data mining is going to help you understand what is relevant, and then it can make good use of that information by assessing it and making predictions based on that information.
3. It can speed up the amount of time that it takes to make informed decisions - decisions that are based on sound data and information.

From here, we need to be able to look at how this data mining is going to work for us. When we look at data mining as a composite discipline, it represents a variety of methods (and techniques) that can be used with a variety of analytic capabilities. And these are all going to come together to address a gamut of organizational needs, ask the right questions, and use varying levels of input or rules from the programmers who design them in order to help us arrive at a good decision.

There are two types of modeling that we can work with here when we do data



mining. These will include descriptive modeling and prescriptive modeling, and each one is going to work in a slightly different way to help us get the results that we would like.

First, let's look at what descriptive modeling means in data mining. This uncovers some of the shared similarities that are found in our historical data in the hopes that we can determine the reasons behind failure or success that happened in your business. This could be something as categorizing your customers by their sentiment or their product preferences. There are a variety of techniques that you are able to use to make this work, and some of them are going to include:

1. **Clustering:** This is when you can group together records that are pretty similar.
2. **Anomaly detection:** This is where we are going to identify some of the outliers and whether they are important or if they can tell you something new.
3. **Association rule learning:** This is when we look at our records and see if there are any relationships that show up between all of them or not.
4. **Principal component analysis:** This is the one we will use when we want to take our variables and determine if there is some kind of relationship that shows up there.
5. **Affinity grouping:** This one is going to be used when we want to group people with common interests, or with goals that are similar. This would be something like people who buy X are often going to buy Y at the same time, and they may also purchase Z at the same time.

Another method that we can work with is known as *predictive modeling*. This one is going to be useful because it is able to go deeper in order to help us classify some events in the future or to help us estimate an outcome that is unknown. This could include something like using credit scoring in order to determine how likely it is that an individual will be able to repay one of their loans or not.

Predictive modeling is going to come into play because it can help us to uncover insights for things like the churn of the customer, the credit defaults, and the response of a certain campaign that you did. Some of the examples of techniques that we can use that fit under the umbrella of predictive modeling

will include the following:

1. *Regression*: This one is going to be the measure of the strength of a relationship between one dependent variable, and a series of variables that are independent.
2. *Neural networks*: This is when we will use programs on our computers in order to detect the patterns that are found in a set of data, make predictions based on that data as well, and to help us learn from all of that data.
3. *Decision trees*: These will be diagrams shaped like trees where each of the branches represent a probable occurrence that could happen at some point.
4. *Support vector machines*: This is going to be an example of a supervised machine learning algorithm that is going to be able to help us learn more about the data we have

The third type of modeling that we can work with is known as *prescriptive modeling*. With the growth of all the data that is unstructured from the web, books, comment fields, audio, PDFs, and email, along with some of the other sources of text, the adoption of a field that is known as text mining is growing quite a bit. It is related to data mining, and it helps us to gain the ability that we need in order to parse successfully, filter, and transform all of the data that is unstructured to help us work with the predictive model to improve the accuracy of any predictions that we make.

In the end of this process, you do not want to look at data mining as a separate or a standalone kind of entity because pre-processing, which is data exploration and data preparation, and post-processing, which will be things like model performance monitoring, scoring, and model validation, and both are going to be just as important as one another.

Prescriptive modeling looks at both the external and the internal variables and constraints in order to recommend one or more courses of action that you could take. For example, it could be used in a way to help your company decide the best marketing offer to send out to each customer, making this more customizable and easier to work with and increasing the likelihood that you are going to get a sale. Some of the techniques that you are going to be able to use that go along with prescriptive modeling include:

1. **Predictive analytics plus rules**: This is when you will spend some time developing if and then rules from patterns and

predicting the outcomes that you want.

**2. Marketing optimization:** This is when we will simulate the most advantageous media mix in real-time in order to increase the amount of return on investment that you are able to get out of every action that you decide to take.

Data mining is going to be such an important thing that you can do when it comes to your predictive analysis. It will ensure that you are able to get ahold of the information that you want and then go through it and mine through it until you find the information that is the most important.

As a company in the modern world, it is likely that you are taking in a ton of information all of the time. And while this is definitely a good thing for your business, it is more about how you use this data, rather than the exact amount or type of data you decide to collect. You can spend all day collecting that data and store it for years to come, but if you don't go through and mine some of that data, and learn the insights that are inside of it, then the data is going to be worthless to you.

Data mining is going to help you to actually get some use out of the data you are working with. Instead of just storing that data on a server or on the cloud, without any idea of what is in it or how you can use it, you can work with data mining in order to get the most out of the data that you collect this can bring you a lot of the important insights and trends that are found in the data, so that you can then, at a later time, go through and complete your predictive analysis. When it comes to getting something out of your data, and making sure that your predictive analysis is going to be successful so you can make good predictions, make sure that you complete a good data mining process on the data ahead of time as well.

## **Chapter 4: Predicting Future Events with Predictive Modeling**

Now that we have had some time to look at data mining and data analytics when it comes to predictive analysis, it is time to move on, taking a look at what is known as predictive modeling. This is slightly different from some of the other topics that we have spent our time on so far, but now it is time to focus on this a bit more and see how it can factor into the predictive analysis that we want to work with.

First, predictive modeling is going to be the process that we are able to use the results that we already know in order to create, process, and then validate a model that we are able to use to forecast future outcomes. It is a tool that we will see in predictive analytics, and it is also part of data mining that helps us to answer the question, “What might possibly happen in the future?”

The whole point of working with predictive modeling and predictive analysis is that we want to be able to look through data and trends and find out what is going to happen next, and the modeling that we are able to do here will help us to get it all done. You must use the right algorithms and the right methods like what we talked about before, but it can help us to make accurate predictions on what is going to happen in the future.

### **Understanding How Predictive Modeling Works**

The rapid migration and change in our world over to digital products has created a sea of data, one that is easily available and accessible to any business who wants to take the time to collect it. Big data is already being utilized by many companies in order to improve the kind of dynamic that is there in the customer to business relationship. This vast amount of data that is collected in real-time, rather than after the fact, is gotten from a variety of sources like cloud computing platforms, cell phone data, internet browsing history, and social media, to name a few.

When we take some time to analyze events that are historical, there is a higher probability that a business is able to predict what would happen in the future, and then they can plan accordingly. Having a machine and some algorithms and models do the work can make this a lot easier. The algorithm can go through the information in no time at all and will ensure that you will see some results and some predictions in no time.

Yes, it is possible for a human to go through this data rather than using the

programming and algorithms. But when you have millions of data points, and more data is added all the time, this becomes a huge undertaking that your employees have to handle, and this is not a good thing at all. It is too big of a task - things are going to get missed - and your employees will not be able to keep up with the work or the information that is streaming in on them. This is why it is best to work with the programming and other options that are available to you.

The data that is coming into the company will often be unstructured and really complex, often too complex for humans to be able to get through and analyze in a short amount of time. Because of all the complexity that enormous amounts of data are going to present, companies are going to turn more towards predictive analytics tools to help them make predictions and to forecast the outcome of an event likely to happen in the near future, and then they can make their plans according to this likelihood.

How does this kind of analysis really work?

Now that we know a bit more about the modeling that comes here, we also need to look at how the predictive analysis - the main topic that we are talking about in this guidebook - is going to work. The predictive analytics will help us collect and then process this data in huge amounts while using powerful computers to help us to assess what has happened in the past. When this part of it is done, it is then able to provide us with a good assessment of what it thinks is going to happen in the future.

The predictive analytics can use predictors, or sometimes known features, in order to create predictive models that we can then use to obtain an output. This model is going to help us learn how these different points of data are going to connect with one another. Two of the most widely uses of predictive modeling techniques include the neural networks and the regression that we talked about before.

### **Other Things to Consider**

There are also a few other topics that we should discuss when it comes to working on a predictive model. In the field of statistics, we will look at regression and see that it talks about the linear relationship between the input and the output variables that we would like to use. So, if a predictive model has a linear function, it is going to require one predictor, or one feature, to help us predict what the outcome, or the output, will be in the end.

A good example of this can be seen in the banking industry. In this example, a bank that hopes to detect when fraud or money laundering is about to

happen, in the early stages before things start to get really bad, may incorporate a linear predictive model to help them with this. The reason for this is because the main goal of the bank is to know which of its customers are likely to engage in the activities that come with money laundering at one point or another.

To help make this kind of model and ensure that it is accurate, all of the data for the customers of the bank will be presented to the model. Then, the predictive model is going to be built around the dollar value of transfers that each customer was making during the time the bank specified in the model.

During this time, the model is going through a process of learning so that it can easily recognize the difference between the transactions that are normal, and the transactions that are going to include money laundering. The optimal outcome that we get out of this model should be a pattern that will signal when this money laundering is happening, and when it didn't. If the model perceives that there is a pattern of fraud emerging in one customer based on the learning they did, it is going to create a signal for action, and then the fraud analysts at the bank will take over and look into it more closely.

The predictive models are also going to be used with neural networks powered by deep learning and machine learning, which are both going to be fields found in artificial intelligence. The neural networks will be inspired by the way that the brain works, and they are simply created with a web of interconnected nodes in hierarchical levels, which are the basic foundation that we find in artificial intelligence.

One of the most important capabilities of neural networks is their ability to handle a non-linear data relationship better than any other available algorithm. These neural networks can create relationships and patterns between the variables that would be either too time consuming, or impossible, for a human analyst to handle.

So, while a bank is able to go through this model and input some of the known variables, such as the value of transfers initiated by their customers into the model to get the desired outcome of who is the most likely to engage in the laundering of money, a neural network is able to create a pattern and relationships between the input variables, like the geographic location of the user, the time logged in, the IP address, the sender and the recipient of the funds, and other features that could be included when it comes to activities of money laundering.

Some of the other techniques that are used in predictive modeling and that

financial institutions like to use, including decision trees, time series data mining, and even the Bayesian analysis. Companies that are going to take advantage of big data through predictive modeling measures are better able to understand how their customers engage with their products and can identify any of the potential opportunities and risks for the company, saving them a lot of time and money in the process.

Working with a predictive model can really help you out a lot here because it ensures that you are able to take all of that big data that you have been collecting, and turn it into something that can make predictions easily for you. And this is one of the biggest goals that comes with predictive analysis. You want to be able to use the information to drive your business decisions, rather than just guessing and hoping your predictions are right. If you are able to use this kind of prediction in the proper manner, you will find that it is easier than ever to know exactly what is going to happen in the future, based off your historical data, and make decisions that can help you beat out the competition. You will see some great results in no time!

## Chapter 5: Using Statistics to Help Predict Future Events

We have spent a bit of time bringing up the idea of statistics a few times throughout this guidebook, but we have not gone into much depth about what this means, how this can work, and why we would want to use statistics to help us out with some of these future events. It may seem like math is not something that we need to bring to the table when we are working here, but it can definitely take us where we want to be and can be the step that is needed to help us make accurate predictions that will propel our business forward.

Before we start with this, we need to take a look at a famous statistician and forecaster, Nat Silver, who has been able to make his own name in the industry by predicting the kind of performance that we are able to see with baseball players. He then went on to start his own blog known as FiveThirtyEight, which is known for very accurate predictions in many of the recent political contests out there.

While this may seem like it is beside the point, it is actually going to help us see how statistics can come into play and maybe the answer that we need in order to why we need to use this field to make predictions in the future. In fact, Silver spends some time talking about a few strategies that we are able to focus on to help us know where we are, and what course we want to take, whether we are talking about our own personal future or the future of our business against the competition.

There are a few issues that we are able to pay attention to that can help us to really see a difference in how well we can predict the future, and these will include:

1. Focus more on what is important. There are two great inventions out there, the internet and the printing press, that have helped us to see an increase in the amount of information that is available to us at a fast rate. But just like with Big Data, we do not want to look at all of the information. Some of it may seem like noise, or it may be silly or false. The challenge here is for a business to use statistics to help them focus on the information that actually matters while leaving all the rest behind.
2. There are a lot of issues, and many of them will end up clouding our vision. It is common for people to attach a lot of importance to



recent events and the newest data in the set because they seem more dramatic than the others, but most of them are not that important. We also focus on the familiar, so it is likely that a business may stray towards staying where they are, rather than moving forward and making changes that will help them.

3. The best thing that you can do to break out of this kind of tunnel vision and then test your views from the real world, acknowledging some of the mistakes that you may have made in the real world, and the things that you are struggling to let go of now, learn from them and then actually follow the information that is out there that is meant to help you see some success.

4. Know what a forecast and a prediction are all about. To start, the prediction is going to be a statement that is specific about when and how something is going to occur. A forecast, on the other hand, is a probabilistic statement. Science is not going to be able to predict with 100 percent accuracy when or where an earthquake will occur. But it will be able to forecast with some confidence that in the next 100 years there is going to be a big earthquake that happens in California.

In a similar manner, you can also make these predictions about your own company. You can figure out what is likely to happen over a period of time, rather than the exact date and time, to help you know the right course of action to take.

With some of these in mind, we need to take some time to look at statistics and what it is all about. Why would we want to know more about statistics and all that comes with it, rather than just using some of the other methods that we have been able to look at through this guidebook? Let's see more about this field and where it comes in when we work on our own predictive analysis along the way.

Statistics is going to be one of the forms of mathematical analysis that you can use, one that is able to use a quantified model, representation, and synopses for a given set of experimental data, and sometimes for studies that are done in real life. Statistics studies the methodologies used to gather, review, analyze, and draw the right conclusions from the data that you have. The good news is that there are a number of measures that you are able to use to make this happen, and these options are going to include:

- The mean
- A regression analysis
- Skewness
- Analysis of variance
- Variance
- Kurtosis

We need to take some time to understand statistics a little bit more. Statistics is going to be a term that is used in order to summarize a process that the analyst is going to use to characterize their set of data. If the set of data you have is going to depend on a sample of a larger population for example, then it is possible for the analyst to develop interpretations about the population, and this is going to be based mostly on the statistical outcomes that you get from that particular sample you did the work on.

When you do an analysis using these kinds of techniques, it is going to involve the process of both gathering and evaluating all of the data that you need, and then you have to perform a summary of the data, but it must be done in a format that is more mathematical.

While the word may make some of us a bit nervous about working with, statistics is used in a lot of different disciplines on a regular basis, and it is not that unusual to find it in places like manufacturing, government, humanities, social sciences, physical sciences, business, and psychology to name a few. The data that is statistical is going to be gathered using a sample procedure in most cases, but it can also use other methods on occasion.

There are many options when working with some of the statistical methods, but the two types that are the most common when it comes to analyzing the data that you have will include the inferential statistics and the descriptive statistics. Descriptive statistics are going to be used to help synopsise data from a sample exercising the mean, or sometimes the standard deviation. Or the other hand, the inferential statistics method is going to be used when you have the data, and it is viewed as a type of subclass that goes with the specific population that you are working with.

### **Different Types of Statistics**

Statistics is a general term that is pretty broad, which means that there are going to be a lot of other models that can fit with it. This is perfect, though, because it allows us to see a lot of different ways that we can predict the future with the use of this term. Some of the different types of statistics that

you are able to work with will include:

The mean. This is simply the mathematical average of a group of numerals, which needs to include at least two numbers or more. The mean for a set of numbers that you are specifying can be computed in many ways, including the arithmetic mean, which shows how well a specific commodity performs over a specified period of time. Then you have a geometric mean, which shows that the performance results of an investor's portfolio invested in that the same commodity over the same period of time.

The next option is regression analysis. This one can determine the extent to which factors you determine (such as the price of your product or the interest rates) are going to influence the fluctuations of the price of the asset. This is shown in the form of a line that is straight and is usually called linear regression.

You can work with a process that is known as the skewness, as well. This one is a term that you may not have heard about much, but it is still an important part that comes in here. Most sets of data, including the commodity returns and the stock prices, are going to either have a skew that is positive, which means that the curve is skewed a bit towards the left of the average of the data or a negative skew that goes the other way.

We can also focus on working with the process of kurtosis. This kurtosis is going to measure whether the data is going to be light-tailed, which means it is less prone to the outliers - or heavy-tailed - which means that it is more prone to the outliers compared to what we will see when we look at the normal distribution.

The sets of data that are high in the kurtosis are going to come with heavy tails, or lots of outliers, which implies that there is a greater risk with the investment in the form of wild returns that happen on occasion. You have to decide what kind of risk you are willing to take with this one before jumping in to see if it is the right option for you.

It is also possible for the kurtosis to go the other way. This is when the set of data has a kurtosis with a light tail, or there is a lack of lots of outliers. This kind of risk for investors is going to be lower than the others, which can keep the investments safer, by may not give you the same returns that you were able to get with the other option.

Finally, we need to look at what is known as a *variance*. This variance is going to be the measurement of the span of numbers in the set of data. This variance is going to measure out the distance each number in the set is from

the mean. Variance can help us out because it allows the investor to determine the amount of risk that they are going to be taking on willingly when they work on a new investment.

As a business, it is important to be able to manage the risk and the reward ratio that you will get when you complete any kind of task or investment. If the risk is higher than the potential reward, it makes sense that you would want to stay away from it because it is more likely that you will lose money, and the amount that you gain is not going to be worth it.

But what happens with those ratios that are in between? There are a lot of areas below the high risk and low reward and figuring out which one is the best for you to take is going to be one of the many challenges that you face as a business owner. Working with statistics is going to be one of the best ways to take a look at this risk to reward ratio that you have, and then can lead you to make the best decisions for your business, whatever this may be in the process. Adding in the right algorithms, and the right kind of statistics to the mix can make this a lot easier to work with.

Using a combination of the methods above, based on what information you are working with - and what you are hoping to get out of that data - can make a world of difference. You will be able to sort through the data that you have and come up with a good idea of what will work and what can help you out the most in the process.

Working with statistics is so important when it comes to a lot of the things that you do with predictive analysis. It can help us to get some of the mathematical parts that we need with the analysis and can ensure that we get the information that we need in the process, making life easier overall. Most of the time, the algorithms that you are going to focus on when doing predictive analysis will use some form of statistics with it, so knowing how this process works can be critical to helping you make some good predictions.

## **Chapter 6: Bringing Data Analysis Into the Mix to Make Predictions**

Another part of the puzzle that we need to bring out here is the idea of data analysis. Just like with some of the other elements that we have spent some time discussing in this guidebook, the data analysis is going to be essential and can help us make some predictions on the data that we bring in. This method may seem similar to some of what we have been able to talk about so far in this guidebook, but it is going to take a slightly different approach, and it is still a good idea for us to take a look at this and see how it can come into the mix to make some reasonable predictions.

To start, we need to take a look at what this data analysis is all about. Data analysis is going to be the practice where we can take some of the raw data that we have and then order and organize it in such a way that it becomes useful, allowing us to cull the important insights and trends that we need from it. It recognizes that with all of this data, there is going to be some good information, but there will also be a lot of noise and a lot of outliers that we need to be careful about, and that we can probably ignore without missing out on too much in the process.

The process of organizing and then thinking about the data that we have is going to be a big key to understanding what the data has inside and what it is not holding onto. There are a lot of ways that someone can come and approach what they see with data analysis, and it is notoriously easy to manipulate the data to say and behave the way that you want during the analysis phase. This allows some people to push the agendas and the conclusions that they want, rather than actually looking at the facts that are presented.

This is why we need to be really careful with the data and that we don't try to manipulate it at all. There is a lot of good insights that are found in the data that we want to look through, but if we start to mess around with the data and not treat it the way that we should, then we are going to make bad decisions that we think are driven by the data, but really are not, and this can harm our business as a whole.

Because of the ability to manipulate and read the data wrong, it is important always to be critical when we look at the information that we are presented. It is important also to pay attention when we have someone presenting data

analysis to us and to think critically about any of the data that we see, and the conclusions that the other person was able to draw.

You may also find, as we have discussed before, that the raw data that we will use in this kind of analysis can come in many forms. This could include any observations that the company took, the responses to surveys that were sent out, and even some measurements. The one that you use, and it is possible to work with more than one if needed, is going to depend on the kind of data that you are trying to work with, and some of your overall goals when you start collecting the data in the first place.

When you have the data in the raw form, you have a ton of data, and often the information is going to be really overwhelming, even if it is really useful for your business. Over the course of the process that you do with data analysis, you will be able to order the raw data in a way that can turn it into something that makes no sense into something that is a lot more useful.

For example, when you first get in the surveys that you send out to your customers, this is going to seem like an extraordinary amount of information that you need to keep track of, and you may feel like you will never get through it to see the insights. But as you go through and tally up the information that you have, you can start to make charts and graphs that can help you see, just at a glance, who answered the survey and what results they were able to give to you along the way.

During your time organizing all of this data, it is likely that you are going to see a variety of trends emerge in the process. These trends are important because you can highlight them in the write up of the data, ensuring that your readers notice the same thing that you do and that they will take this kind of thing into consideration.

Let's say that you were doing a kind of causal survey of preferences when it comes to the type of ice cream someone likes. In this survey, after you went through and asked a big group of people and tallied the information, you found that more women compared to men had a fondness for chocolate. Depending on what your goals are with this kind of survey, this point may be of interest to the person conducting the survey. Being able to model the data with the help of many tools, including mathematics, can sometimes help us to exaggerate these points of interest that show up in the data, which is always going to make it easier for the researcher to see.

There are a lot of different points that are going to show up in the data analysis that you do. Some examples of these are going to include a textual

writeup, graphs, and charts. These methods were designed in order to refine and then distill the data so that the readers were able to get some of the interesting information from all of that data, without having to take all of that large set of data and sort through it on their own.

You will find during this process that summarizing the data is going to be so critical to the supporting arguments and can ensure that you can maintain some of the arguments that you have as well. It also helps us when we want to present the data in a manner that is understandable and clear. This makes it easier for those who are responsible for hearing about the information to know what is inside and can make it helpful when it is time to make some important decisions based on that information.

There are a variety of methods that you are able to use in order to help summarize the data and make sure that the person who needs to see it will be able to understand what is there. Using the charts and graphs are often the best option here, but many times a data scientist is going to add in the raw data that they have in an appendix so that others are able to look up the specifics and double-check the work that they did to build more trust in the process.

When many data scientists encounter the data and the conclusions that are summarized, it is likely that they are going to take them at face value. This is not always the best idea, though, because it is important to know how these conclusions were formed, and whether it is actually something that is accurate or not.

Everyone who sees this data needs to be able to take a look at it in a critical manner. Asking where the data is from can be really important, as is asking about the method that was used for sampling to collect the data, and the size of the sample that was used in the process. It is the source of the data that appears to have a conflict of interest with the type of data that you are being gathered, and this can call the results into question if you are not happy with the conflict of interest that is there.

Researchers who are reputable and trying to actually come up with good and accurate results will make sure that these conflict of interests are not present and will be able to provide the information about all of their data gathering techniques that are used, the source of the funding that helped them to get the work done, and the point of why they collected the data in the beginning so that those who are looking at the data are able to think through whether this is a source of the data that you would like to use.

The good news is that many of the options you use for the data analysis that are used, such as government sources and many of the independent options, are going to be good to work with, and you won't run into this issue. But it is always a good idea to go through and double-check the sources of all the data you are looking for to make sure there isn't a conflict of interest, and to make sure that the data actually supports the conclusions that you are seeing. This will ensure that the data you are using, no matter what you are using it for, will actually be useful for what you would like.

Data analysis is going to be an important section when we work on this process. It is going to ensure that we will see some good results in the process, and makes it easier for us to really take the raw data that we have, and put it into a form that is the most useful to us. Learning how to analyze the data that you have after it is gathered, and even working with some charts and other visualizations can be so important to get the most out of this type of analysis while making some good predictions for your company based on the data.



## **Chapter 7: Using Machine Learning to Make the Predictions For You**

The next topic that we need to spend some time learning about is known as machine learning. Machine learning is a great process to work with because it can take a lot of the analysis that we have already talked about in this guidebook, and will provide the actual power that helps us come up with the predictions that we use in our decision making. You will find that without these algorithms, you may have to go through manually and look through all of this data, and that is wasteful and time-consuming. And with machine learning, there are a lot of choices in the algorithms that you want to use, which can make your life easier to handle any kind of data that you would like.

Before we dive into how machine learning can be used to make predictions, we first need to take a look at what machine learning is all about. Machine learning is going to be a technique that you can use in data analytics that will teach the computer to do what seems to come pretty naturally to animals and humans. What we mean by this is that we can use various techniques to teach a computer how to learn from past experiences.

How does this work? Machine learning works with algorithms that will use a variety of computational methods to learn information directly from data without relying on an equation that is predetermined as a model. The algorithm can adaptively learn how to improve their performance as the number of samples available for learning increases.

For example, your data set could be limited in the beginning because you don't have as much information available. You still use that to do the training and the testing to get the algorithm up and running. After that time, as you add in more data, and the algorithm can learn along the way, you will find that the algorithm gets stronger and will be able to make more accurate predictions as well.

Machine learning is going to be a method of data analysis that is able to automate the process of building analytical models. It is also a branch of artificial intelligence that is going to be based on the whole idea that a system is able to learn from the data it is presented, it can identify the patterns that are there, and it is even able to make its own decisions without a lot of intervention from humans in the process.

Because of all the new computing technologies that are out there, machine learning - as we know it today - is not really the same as the machine learning that we see in the past. It was born out of a form recognition for patterns and the idea of how a computer is able to actually learn, without a programmer there, to ensure it performs a task specifically. Researchers who were interested in some of the things that we are able to do with artificial intelligence also wanted to see if their machines were able to learn from data that it was fed.

The iterative aspect that comes with this machine learning should be seen as an important programming tool because as we expose any of the models we create from this learning to new data, the model is then able to adapt on their own and independently. The machine is going to be able to learn what has happened to it in the past, and the examples it was given, in order to make accurate and reliable predictions in the future.

Thanks to the rise that we see with big data, machine learning is starting to become one of the big techniques that are there to help with a lot of different areas of solving problems. Some of the biggest areas that are going to be able to use machine learning to help solve their problems will include:

1. Computational finance This is when the financial world can use machine learning to help with things like algorithmic trading and credit scoring, to name a few processes.
2. Computer vision and image processing. This is when we can use machine learning to help with processes like object detection, motion detection, and facial recognition.
3. Computational biology. This is when we will use machine learning to help with processes like DNA sequences, drug discovery, and tumor detection.
4. Energy production. This is when we use machine learning to help with price and load forecasting.
5. Manufacturing, aerospace, and automotive. This is when we are going to use machine learning to help out with processes that include predictive maintenance.
6. Natural language processing. This is when we will use the idea of machine learning to help with applications of voice recognition, like what we see on many smartphones.

Machine learning algorithms are helpful because they are going to find some

of the natural patterns that are already in the large amount of data a company holds onto, and then it can generate the right insights while helping you to make good predictions and decisions in the process. These algorithms are going to be used each day in order to make critical decisions in things like energy load forecasting, stock trading, and medical diagnosis, to name a few. A good example of this is with the media. Many media sites are already relying on the algorithms that come with machine learning in order to sift through the millions of options to provide the movie or song recommendations for you. Retailers have been using these algorithms in order to gain some good insights on their customers and the behavior for purchasing that they see there, and then they can provide good recommendations and make changes that improve their bottom line.

We also need to consider some of the times when machine learning can be helpful. You will need to bring out machine learning, and some of the algorithms that come with it, any time that you would like to work with a task that is more complex, or you would like to solve a problem that needs a large amount of data, and there are quite a few variables that go with it, but there isn't an equation or a formula that exists to handle this.

For example, some of the times when you will want to work with machine learning to help your business to grow will include some of the following:

1. Looking at handwritten rules and equations that are going to be too hard to do manually. This could include some of the processes, like speech recognition or facial recognition.
2. When the rules of the task you want to complete are going to change on you all the time; this might include something like transaction records or fraud detection.
3. When the nature of the data that you would like to use keeps changing, and you want to make sure that the program you are using will be able to adapt. This is going to include some processes like predicting the trends in shopping, forecasting energy demand, and automated trading.

Machine learning is going to work with three types of techniques or methods that you want to work with based on the kind of data you want to work with. The types of machine learning that you are able to work with will include supervised machine learning and unsupervised machine learning. Let's dive into both of these and see how they can work to our benefit.

The first type of machine learning that we are able to take a look at is known

as supervised learning. This type of learning is going to be able to build up a model we can use that can make good predictions based on evidence that we have, in the presence of uncertainty. The algorithm that fits this mold is going to be able to take a known set of input data, along with the responses that we already know about the data or the output. It can then train the model that you have to generate some reasonable predictions when it receives new data along the way.

You are going to work with supervised machine learning if you have the known data for the kind of output that you would like to predict. This process involves you providing the right examples to the model so that it can learn. You provide the input, and the output learns from this, and then it can make good predictions based on what it has learned as time goes on, and it receives some new information or input.

There are two types of techniques that are going to work when you want to use supervised machine learning in order to develop your own predictive models. The first one is going to be a classification technique. This kind is going to help us to predict a discrete response. Some examples of this are when you want the model to estimate whether the email is spam or a genuine email. Classification models are going to be able to classify the data that is used as inputs into different categories. Typical applications are going to include things like credit scoring, speech recognition, and medical imaging.

You can work with classification any time that you can take the data and separate it, categorize it, and tag it into specific classes or groups. For example, applications that are going to use handwriting recognition is going to use classification in order to help the program recognize different numbers and letters. When we work with things like image segmentation or object detection, you are going to be working more with the unsupervised machine learning that we will look at in a bit.

There are a variety of different types of algorithms that a data scientist is able to work with when it comes to doing supervised learning on classification problems. Some of the types of algorithms that fit into this will include neural networks, logistic regression, Naïve Bayes, K-nearest neighbors, decision trees, and support vector machines.

Likewise, we can work with the regression techniques. These can predict responses that are continuous, like a change in the temperature of a certain area, or the fluctuations that show up in the demand for power. Some of the typical applications of the regression techniques are going to include

algorithmic trading and electricity load forecasting, to name a few.

You will want to work with some of these regression techniques if you are working with a data range or if the nature of the response that you would like to get out of that data is going to be a real number, such as a time unit failure on your equipment, or the temperature. Some of the regression algorithms that you are likely to work with could be stepwise regression, nonlinear and linear model, and adaptive neuro-fuzzy learning.

The next type of machine learning that we need to focus on for a few minutes is the idea of unsupervised machine learning. This one is not going to need the examples in order to make things work. Instead, it is going to work on its own in order to find the hidden structures that are found in the data. You can use this kind of learning to figure out what inferences are in the set of data consisting of input data that doesn't come with labeled responses.

The most common technique that you can use with unsupervised machine learning is clustering. This clustering is going to be used to help with data analysis that is more exploratory in nature. This means that the clustering is going to help us to find groupings or hidden patterns that are in the data some of the applications that we can see with a cluster analysis will include object recognition, market research, and gene sequence analysis.

A good example of this can be seen with some of the work that a cell phone company partakes in. Let's say that this company would like to optimize the areas where they build up new phone towers. To do this, they can work with unsupervised machine learning in order to estimate the number of clusters of people who rely on these towers, and where these individuals are residing. They can't put up a tower where one or two customers are located, but if they see that there is a significant number of people in an area, it may be a good idea to place a tower there.

A phone can only talk to one of the towers at a time, so the team is going to use the algorithms of clustering to design the best placement of the cell towers, optimizing the reception of the signal for the groups, or in this case the clusters of their customers.

Just like with supervised machine learning, there are also a few different types of clustering that you are able to work with to complete your work with unsupervised machine learning. Some of the options that we can work with include self-organizing maps, the hidden Markov models, Gaussian mixture models, and hierarchical clustering.

The third type of machine learning that you can work with is known as

*reinforcement machine learning*. This one is often one that is forgotten or combined with unsupervised machine learning because the two are so similar. Rather than relying on unlabeled data on its own, reinforcement learning focuses on the idea of trial and error. As it makes errors, it will learn what not to do and will make some changes. If it makes the right decisions, it is going to form some strong connections that will help it to do better the next time it encounters the same example.

### **How can I Decide Which is the Correct Algorithm to Use?**

The next thing to focus on is how you can choose the right algorithm that you can work with. This can be an overwhelming thing to work with. There are going to be a lot of different algorithms that you can work with here to make predictions, both with supervised and unsupervised machine learning, and each of them will take on a different approach when it comes to learning.

There is no best method to work with, and no one size fits all. Finding the right algorithm in some cases is going to be all about trial and error. And even those who have been doing this kind of work for a longer period of time can't always tell whether or not one algorithm is the best one, at least until they try it out.

Often the selection that you make with picking out the right algorithm is going to depend on the type and the size of the data that you are working with, the insights that you are trying to get out of that data, and how you would like to use the insights that you receive. Some of the basic things that you can remember and consider when you are working with these algorithms, and when you are stuck trying to decide between supervised and unsupervised machine learning will include:

1. You want to go with supervised learning at any time that it is best to train your model to make a prediction. You could use it to help you work with a future value that is a continuous variable such as the stock price or a temperature. It can also work with classification, such as identifying the make of the car from footage found in a webcam.
2. You want to go with unsupervised machine learning if you would like to explore your data, and you are looking to train the model in order to find a good representation that is internal, such as taking all of the data that you brought in and then splitting that data into some clusters.

### **How Machine Learning Can Assist in Making Predictions**

Many of the different types of algorithms that come with machine learning can come into play and will help you to make some smart decisions in the process. These algorithms are set up to handle a lot of the information and the data that you are collecting and can make predictions for you. In fact, this is one of the main things that machine learning is able to do!

With machine learning, we already recognize that it is really hard for us to go through and sort through all of the data that we collect for the business. When millions of data points come at us all of the time, from social media, a website, survey results, or something else, and this is just too much for a single person to go through, or even a team to go through. It would take too long, and since there is a constant stream of this data, it is impossible for them to come up with relevant information and decisions in a timely manner. You can certainly try it, but by the time the person actually went through all of that information and saw the trends or the patterns, the information would be old, and they would be way behind the competition. It is much better to make your decisions based on intuition and business sense and see what results you get, rather than spending time manually going through all the data that is available.

Rather than doing this, you can work with some of the algorithms that come with machine learning to look through the information. These algorithms are able to take that information and go through it quickly, providing you with a good prediction and all the possible outcomes that can happen as well based on the data that you feed in. While it could take years for an individual to attempt to keep up with this information and see the trends, the algorithm can get this done really quickly, and could keep up with the constant stream that the company gets with the data.

There are many different types of machine learning that you can work with and a bunch of algorithms that fit into the three types that we talked about earlier including supervised, unsupervised, and reinforcement learning. The type of algorithm that you decide to use when it comes to machine learning will make a big difference in the results that you see, and often it depends on the kind of data that you are using, and the end results that you are hoping to get out of it in the process.

If you work with machine learning in the proper manner, you will be able to set up a model with the chosen algorithm that will then be able to handle all of the data as it comes in. Some algorithms can take in a steady stream of data as it comes in, and others need you to manually go in and provide the

data at the times when you want the new predictions. You can then set up the model so that it handles all your data in a consistent manner while dealing with some of the other types of analysis that we have talked about in this guidebook already.

Most of the industries that are out there that are already working with large amounts of data are going to be able to recognize the kind of value that they would get with using the technology that comes with machine learning. By being able to actually get through this data and glean some good insights from it, and being able to do this close to real-time, the company is then able to work in a more efficient manner in order to gain a big advantage of others in their same industry.

And this is the beauty of working with machine learning. We are able to do things that may have seemed impossible in the past are possible now with the help of machine learning. Businesses that are handling more data than ever before are finding the value of working with machine learning to help them get their work done. They can get through this information faster than would be possible with a person looking through it on their own, and can give them that competitive edge over others.

There are a lot of different companies that will be able to benefit from a program that can run on machine learning. Some of the different industries that are already using this kind of technology will include financial services, government, health care, retail, oil and gas, transportation, and more.

### **The Prediction Explanation**

Traditionally, when we look at some of the models found in machine learning, they are not going to provide us insights into why or how the model arrived at the outcome that it did. This is going to make it a bit harder to explain in an objective manner the decisions that are made, and the actions that it took based on these models. Prediction explanations are going to avoid the black box syndrome by describing which characteristics or feature variables, are going to have the biggest impact on the outcome that you will see with the outcomes of the models.

When we need to know the reasons that are behind the outcome we see with a model, and this background is as important as the outcomes themselves, the prediction explanations are going to come in and uncover some of the factors that can come and contribute to the outcomes. We may not get the algorithm to come in and tell us exactly what it is doing all of the time, but we can get it to help us see some of the most probable steps that were taken in the process.



It can be frustrating not to know the real reasons why an algorithm is going to do one task over another, or how it came up with the prediction or the outcome that it presents to us. This is not good for business either. How are we supposed to know if the result that we got is the real one that we need, or if it is one that we need to be worried about because some outliers got in the way? Working with the prediction explanation, along with the algorithm that we have, can make a world of difference.

For example, banks are already using models to help them determine whether it is a good idea to approve a loan for a particular customer or not. These financial institutions are able to use what is known as prediction explanations in order to gain some insight into why an application was rejected or accepted. Then the loan officer would be able to see whether the reasons make sense or if something is going wrong with the information.

With the insight that the prediction explanation provides, it is easier for the bank to develop some models that will help them comply with any of the regulations they have. They can also come up with a model that can explain the outcomes of the model to their stakeholders, and a model that will identify some of the high impact factors to help the bank stay on track with some of their business strategies.

Machine learning can be one of the best topics that you are able to use in order to get the algorithms that come with it to work well and to ensure that you are going to get a good idea of how the prediction happened in the first place. This may seem like something that isn't as big of a deal, but before we base a decision on what the algorithm shows to us, we need to see how it is showing us the predictions and how it comes up with the conclusion that it presented to you as well. The prediction explanation is going to help you to make this all work and will ensure that you are actually making smart and accurate decisions based on your data.

## **Chapter 8: How to Avoid Prediction Traps, Avoid Bias, and Make the Best Predictions**

As we have explored in this guidebook, there are a lot of things that you are able to do with predictive analysis, and you may be surprised at all of the different methods that you can bring on board in order to see some amazing results. While it is important to learn as much about the benefits of this predictive analysis as possible, and we do go in-depth about this more in the next chapter, there are a few things that we need to be wary of, and learn how to watch out for, if we truly want to make predictive models that are able to help us out.

While many companies are ready to jump on board with the ideas of the predictive analysis because they like all of the opportunities that are presented with it, and all that this is going to be able to help them out with, we do need to focus a bit on being careful about some of the traps and the biases that can sneak into our models. Sometimes we know those biases and traps are there because we set them to get the answers that we want. If this is something that you plan to do from the start, then just stop now and don't do any more work. That will just be a waste of your time, energy, and resources because you already know the answer you would like to get.

Often though, the traps and biases are going to sneak their way into the model, and we didn't even realize what was going on in the first place. This can be just as dangerous, even though we don't realize what is going on, because we are allowing our data to be tainted with, and this can harm the end result that we end up getting.

Our goal here is actually to learn what is inside of all that data. Approaching it with a clear and open mind, rather than focusing on what we think should be there, is the key to helping us actually learn something new. And you may find that during this practice, you are going to find a lot of useful information and insights that you were not expecting to find in the first place. That is part of the fun and the excitement of the predictive analysis. It allows you to learn something new, which can go a long way in ensuring that you see some good results for your business.

With this said, we now need to take some time to use our predictive analytics, but first, we have to focus on some of the ways that we can avoid prediction traps and any of the biases that we may try to get into the models so that we

can look at the information with new eyes, and glean some useful information out of it in no time at all.

### **How to Avoid some of the Traps and Biases**

You will find that predictive analytics can yield amazing results. The lift that you can see in your business is going to be achieved when you base some of your future decisions on the observed patterns that are found in historical data, and we find that this can far outweigh anything that can be achieved when we rely on anecdotal events or our gut feelings. These decisions are going to be made using evidence and facts to back them up, which can ensure that you will see the best results.

There are a lot of examples that will show us the possible lift that we are able to see through all of the different industries, but one test that was done recently that is done in the retail sector will show that applying stable predictive models was able to give a huge increase, up to fivefold, in the take up of a product when we were comparing it to a random sample. And if this weren't enough to convince us that this was a good thing to spend our time on, we have to admit that there wouldn't be as much focus on predictive analysis, and some of the cool things that we can do with it and machine learning, if it wasn't helping businesses in many different manners, and if it was not providing these businesses with some good results.

Of course, we can't assume that these predictive models are going to be accurate all of the time, and we have to take the necessary precautions to make sure that we are using them in the proper manner. They can be sensitive to change sometimes, and they have a propensity to make sudden shifts up and down based on the results that we put in and the kind of data that we are going to use for the model. Having someone who is able to handle the information and the model, and who will use it in the proper manner is going to be imperative to make this process work.

The commoditizing of machine learning is going to be useful because it helps data science become more accessible to the non-data scientists of the world than we have seen in the past. But with this in mind, we have to make sure to avoid some of the common pitfalls that come with predictive analytics, in order to keep the models performing as expected.

We like to think in many cases that the predictive model is going to be able to solve all of our problems. And sometimes, we use it to prove the hypothesis that we already have in our minds. These are just two of the pitfalls that come with this kind of analysis, and it can really cause some issues with the model

you are using. Learning what these pitfalls are all about, and how to avoid them can make a big difference in the accuracy and efficiency that you will see from these models. Some of the most common pitfalls that we may see with our predictive models, and that we need to watch out for includes:

The first bias and trap that we need to watch out for are making some assumptions that are incorrect about the underlying data we use for training. Rushing into this process and making assumptions on the training data may seem to save us some time, but it is going to lead to some troubles. As a data analyst and someone who really wants to get at the root of the information found in your data, it is much better to take some time to understand the data and then find the trends, the distributions, the outliers, and the missing values.

It is also possible that the data analyst is going to try and work with low volumes. This may seem like the best option to work with a new model, but it is going to harm your model and will not make it as strong as you want. Low volumes are actually meant to be your unhappy place in this process. These low volumes are going to lead to models that are unreliable, unstable, and statistically weak.

This means that you need to take some time to gather and clean off more data before you work on your models. This ensures that you get some strong models out of the process because the training data that you are working with will make it easier to get a strong and reliable model that you are able to use to make predictions.

Next on the list is going to be the trap of overfitting things. What this means is that you are creating a new model that has a lot of branches. This may seem like a good idea to have more options and more branches because it seems to provide us with a lot of better discrimination when it comes to the target variable. But when we bring this over to the real world, we will find that all that it does is introduce too much noise into our model, and makes the model less reliable and accurate in the process.

The training data is something that we need to spend time on, and we have to make sure that we are eliminating this training bias as much as possible. For example, you only took the time to offer a certain product to the millennials in your target audience. With this, you will find that the millennials are going to come in really strongly on this model. Since they did not get a choice to work with, it was hard to figure out what they would actually behave like if they had some choices.

We also need to look at the test data and the training data. One of the biggest mistakes that can happen when you work with both is that you add in the test data with your *training data*. This has led to a few fails because it gives the impression that your model is the best one out there, and it is going to perform fantastically. But in reality, when you do this, you will find that the model you want to work with does not work all that well, and it is actually a broken model.

In the world of predictive analytics, if the results you receive seem to be too good to be true, then it is often a good idea to go back through and see what this is all about. Spending some more time and money to validate the results you get when they seem too good to be true and even getting another opinion to look over your work to see what mistakes may have been made will ensure that you get a model that is going actually to work for you.

Companies who want to work with predictive analysis also need to learn how to be creative with the data that they are provided. If the creativity is kept out of the mix, then this is going to be another trap that we fall into that is going to harm the model we want to focus on. You may find that the predictive model that you are working with is going to be improved quite a bit when you can add in a few clever features or characteristics that are going to explain, in a much better and clearer method, some of the trends that show up in the data.

There are too many times that a data scientist will play by the book, and won't work with any sort of creativity in the process. They will just take the data that they are provided, not spending enough time considering some of the features they could add which are more creative. This creativity is going to help add some more strength to the models you are creating, adding in the strength in ways that even improving the algorithm will not be able to achieve all on its own either.

We also have to remember that we can't expect our machines to understand business. There are a lot of cool things that we are able to do with our machines, and when we add in artificial intelligence and machine learning to the mix, we can teach our computers to handle a lot of the same things that humans can. But we have to understand that right now at least; our machines are not able to understand business or how to handle it.

Machines - as of right now - are not able to figure out what the business problem is and how it can best tackle the problem for you. This is not always something that is easy to understand, and we often need to give it some

careful thought. Businesses spend a lot of time talking out decisions and researching it before making any decisions, and this is not something that we can teach a computer how to do right now. Expecting that the machine is going to come in and handle all the problems that we have, without any interactions from us at all, is going to set us up for failure.

Another thing that we need to consider here is that we don't want to use the wrong metric to measure out the performance of the model. Let's say as an example that out of 10,000 cases, you only have two fraudulent cases, and the rest are just fine. If the performance metric that was used in the training of the model was completely straightforward accuracy, the model is going to attempt to maximize the accuracy that it sees.

Thus, if the model is going to predict that all the 10,000 cases are not fraud, then the model would come in with a 99.8 percent accuracy. This may seem really great, but it is not going to serve us much purpose when it comes to seeing whether or not some of the cases had fraud or not. It simply means that it identified 99.98 percent of the non-fraud instances in the correct manner for the company.

When you are working with rare event modeling, of which the fraud example above is going to help us out, it is best if you can skip the accuracy metric that we used, and instead go with an alternative approach to help us figure out what is going on here.

### **Using plain linear models when the interaction needs to be non-linear.**

This is most likely to happen when you would like to build up a binary classifier, and you choose the logistic regression for this method - when, the relationship that shows up between the features is not going to be linear. In the case above, working with a model that is tree-based or the support vector machines will work better.

This is where it is helpful to know more about the algorithms that you want to use, so you can make sure you are not using the wrong ones on your data. Not knowing which methods can be applicable to the problem you are trying to handle can result in models that are not very good, and the predictions that you get out of them will not be accurate either.

Next on the list is going to be the outliers. These outliers are usually going to deserve at least a little bit of your attention, or they should be ignored. There are going to be a few of these models that are going to be really sensitive to the outliers, and if you use these algorithms, and you forget to take them out, or at least cater to them, then this leads to some bad performance in the model

that you want to work with.

You also need to be careful about regularizations without any standardizations. Many of the companies who choose to work with this predictive analysis are not going to be aware of some of the redundancy of applying regularization to the features of the model, without first going through and standardizing the data to ensure that the data is on the same scale.

If you choose to do this, then the regularization would end up as biased because it is going to penalize the features that are on a smaller scale that you want to use here. This could be a bad thing because it kicks out some of the information that you need, and can make it hard to really see the information and the predictions that you need in the process.

As a company, you need to make sure that you take into account the real-time scoring environment that is present. Sometimes a company using the predictive model is going to get distracted by building the model that is the most perfect. They want everything to be in place and ready to go at the right times (and the right places) all the time.

This is a great goal to have, but it is not realistic. When it comes to deployment, when we work on a model that we want to make perfect, this often becomes so complex that the model can't be integrated into a system that is operational. You do not want to spend so much time on a model, working to make sure that everything is perfect, just to find out that it won't work, and you need to make changes again.

We also need to not fall into the trap of using some characteristics that will not be there in the future, usually because of some operational reasons. One thing that the company may identify is some predictive characteristics, like gender, but because of regulations, you can't use this kind of field to help out with modeling. You also will find that it can't be used in capturing the field that has been suspended and will be available in the future for use in the model.

The final trap that we are going to take a look at is not considering some of the implications of the real world with the model, and some of the possible fallout that could happen when we do one of these analytics. One example of this is the retailer Target. They made headlines a few years ago when a reporter from the New York Times brought it to the attention of the public that the analytics models that Target used were able to predict the pregnancy of a teenager based on how they were shopping on the store.

What this can tell us is just because we can find out some information, doesn't necessarily mean that we should. Yes, there was probably enough data information that was going through their system that they could find out this information based on the age of the customer and what they were shopping for, such as a pregnancy test, but did they really need to know this? Most likely not. It is important that you spend some time looking at the right information, the information that you need to know to grow your business, rather than just collecting information because you can.

There is a lot of useful information that someone can glean when they decide to use predictive analysis in their business. These companies are already gathering a large amount of information on a regular basis already, so why not use it to provide better customer service, and to beat out the competition. But we do need to remember that taking care of the information that we use, and not falling into the traps and biases that come with that information can be important as well.



## **Chapter 9: Top Reasons to Implement predictive analysis**

We have spent some time talking about the various benefits that come with using one of these predictive analytics in your business, but we need to take this a bit further to see how things will go when you actually implement it into your business model. There are a lot of different options when it comes to the way that you will use the predictive analysis, and we have spent some time discussing many of them as we worked through this guidebook. With this in mind, it is time for us to look more in-depth at why you and your business should be using these predictive analytics in the first place.

First, let's take a look at the world of marketing. Marketers are always thinking about new and more efficient ways that they can engage with their customers. According to an insights case study from Forbes called "How Predictive Analytics Helps Cox Communication Tune into Customers," employing the predictive analytics to gain a deeper insight into some of the trends with the consumer has proven to be more successful for many companies, not only for Cox Communication, but for other options as well.

The telecommunications industry is going to be prone to a higher amount of turnover from their customers, mainly because the costs for switching are slim to none, and often it ends up saving them a lot of money. Companies that are in this industry are going to look for new ways to differentiate themselves from their competitors in the hopes of retaining some new customers. It is possible that if predictive analytics is used properly, it is going to help these companies to find the right solutions that will allow them to understand better, and then retain their customers. It can even help them to bring in a new customer in a more effective manner.

It is believed that there are five top reasons why a company would want to use these predictive analytics to help them out in different areas. The top reasons that come with predictive analysis will be predicting trends, understanding the customer, drive decision making that is more strategic, to improve the performance of the business, and to predict behavior. Let's take a closer look at how these can all be done and why a business would want to spend their time on each of these.

### **Predicting Trends that are Coming Soon**

One of the first things that businesses are going to do when they work on predictive analysis is figuring out some of the upcoming trends that are in the

market. The companies that are able to learn about the trends first, and stick with these trends, are the ones that will be able to beat out the competition and actually gain a new corner of the market.

Wouldn't you rather be the company that comes out in a brand new market, or comes up with the newest product that everyone wants, first? Or do you want to watch other companies do these things, and then try to follow after them, hoping to get some form of the market in the process? Using the predictive analysis is one of the ways that you can ensure that you will be ahead of the competition, rather than trying to play catchup and trying to get some of the markets that is leftover.

Let's look at an example. Amazon came out several years ago with Kindle, a device that would allow you to read any book from their library - anywhere you wanted to go. This made it easier for customers to have an easy way to take books around with them, and they could easily sit and enjoy the books without having to carry a whole pile along with them. This was a revolutionary idea, one of a kind, and it is likely that Amazon used something like predictive analysis to help come up with this idea, and found that their customers would really enjoy a product like this.

Amazon got to reap the rewards of this and was one of the first to release this to the market. Everyone else had to try and catch up in order to get obtain some market share. While it is true that there are other readers out there, Amazon is still considered number one with this, and the other options had to really struggle to make theirs unique and different in order to stand out and get some of the customers.

Your business can benefit from doing this kind of thing as well; you just need to learn how to collect the right data, and then perform the right kind of analysis on it to see what your customers want, and what some of the new trends will be. If you can spot these trends and understand what the customer wants before anyone else in the industry does, you can easily step in and really make a product that is first of its kind.

Sometimes, the product is not going to be as big as the Kindle Reader, and maybe it won't change up the whole industry, but it could help you to get a new market for your company, or reach a new demographics before others do. Any kind of good trend that you can jump on first that gets you ahead of the competition is going to do wonders for helping your company to grow and to thrive in the long run.

**Add some Artificial Intelligence to the Mix**

Artificial intelligence, often referred to as “AI,” is an area of computer science. Which can emphasize the creation of systems and intelligent machines and can work - and often react - like what we see with humans. A lot of different computer activities use AI, and are designed to help with planning, learning, problem-solving, and speech recognition.

Basically, when we are working with artificial intelligence, we ~~are going to~~ work with a branch of the larger field of computer science. In this particular part, we are looking at a process that can create intelligent machines, ones that are able to think and act on their own. And because of the power that comes with these systems, it has easily become an essential part of the technology industry as more and more businesses and companies work to use this along the way.

Research that has been done on AI is specialized and highly technical. This could be due to the nature of how artificial intelligence works and all of the different parts that come with it. Some of the core problems that come with the AI include programming the computer for a certain trait. Including the ability to move and manipulate a variety of objects, learning, perception, planning, knowledge, problem-solving, and reasoning.

Knowledge will be power when it comes to working with artificial intelligence, and this knowledge engineering is going to be part of the core of research in this field. Machines are often able to learn how to act and react as a human does, but only if they have been given enough information about how the world works and how they should behave.

The next thing that we need to explore here is why artificial intelligence is so important. Why do we need to spend so much time going through some of the basics that come with artificial intelligence, and why is this a topic that matters at all? Artificial intelligence is starting to take over the world, and learning how to use it for needs can make a big difference in how well companies perform for the customers, and so much more. There are a number of reasons why artificial intelligence would be an important thing to learn how to do. These reasons are going to include:

AI is able to automate some of the repetitive learning that goes on and can help lead to more discoveries through data. AI is going to be different than what we see with hardware-driven robotic automation. Instead of having to go through and automating some of the manual tasks, AI is going to work on frequent, computerized, and high volume tasks in a reliable manner and without getting tired. To get this particular type of automation going, human

inquiry is still going to be essential when it comes to getting that system set up in the proper manner and asking the questions that are needed.

In addition, AI is going to make sure that a product you already have is going to have more intelligence added to it. In most cases, AI is not going to be sold with the help of an individual application. Instead, the products that you are already using will be improved with the help of AI. Many different smart machines can be combined together with all of the data the company holds onto to make sure that the technology is improved, no matter which field it is in.

AI is going to be able to adapt through the progressive learning algorithms so that the data gets to do the programming. AI is able to find any of the regularities or the structure in data so that the algorithm is able to gain a new skill. This algorithm is going to turn into either a predictor or a classifier. So, just like we see the algorithm being able to teach itself the right steps to playing chess, it is also able to use this kind of idea in order to recommend which product the customer should purchase next.

### **Visualizations Can Make It Easier to Show Results to Others**

While a data scientist is going to spend their time going through all of that data, and trying to come up with some of the patterns and trends that are found inside, it is going to be the business owners and the shareholders who really need to know more about the information and can use that data to make predictions. The data scientist can make all of the algorithms and models that they want, but if no one outside of that field is able to understand the information, then it pretty much becomes useless.

Many data scientists need to not only gather the data and then find the trends, but they must make sure that they can take that information and put it into some kind of graphic form to see results. There are a few different methods and models that you are able to use to make sure that this visualization works, such as a chart, a pie chart, a histogram, or a line chart, but you need to make sure that it matches up with the information that you would like to show off.

Sure, you could just present the findings and all of the numbers to others, but this is not the most effective way for them to understand what is inside. These visualizations will allow anyone who wants to, to just look at the information and the findings in a glance, rather than having to read through it all and try to figure out what is there.

Picking the right visualization to go with the work that you are trying to do can be imperative. It will ensure that you are actually representing the data in

the right manner, rather than making things worse. You can experiment with a few different options that fit here in order to see which one seems to represent the data that you have the best.

Of course, it is fine to add in the actual information and the predictions that were made to create those visuals. This allows anyone who would like to check your information and data to go back through it and give you some results. But the visual is going to be helpful because it gives us all the information at a glance, and the predictive analysis - if it is done in the proper manner - can do this for us as well.

### **Drive Decisions that are Better for the Company**

Some of the best decisions that you can make for your company are going to come after performing predictive analysis. These decisions are going to be based on facts rather than just a guess, even if it is an educated guess. Sometimes, the lack of knowledge that we have is surprising, and this is going to harm us if we are not careful about what we are doing.

In the past, there wasn't such a thing as big data or any of the analysis that we have talked about so far in this guidebook. We were not able to pull up data from pretty much any source and then use that information to guide the decisions that we made. Instead, people would have to ask for opinions from others and use some of their own knowledge of the business to determine the right course of action to take.

Sometimes, the business owner would be good at what they were doing, and they would make predictions that we're able to move their business forward. And the longer the individual was in business, the better chance they had of making good predictions as well. But even those who have been in the business for a long time may find that they can fail and run into trouble if they are not careful, and without the proper data behind them to help guide those decisions, it is possible for anyone to fail.

With this in mind, using predictive analysis is going to help with these decisions. Whether you have been in the business for a long period of time, or you are brand new to the industry, these analysis give you the hard facts and data that you need to really see what is going on with your customers, with your business, and even with the industry, so you can make the decisions that are right for your needs.

With this analysis, you have the facts behind you. And often, there are a lot of facts that show up thanks to all the collection methods and the storage solutions for gaining and holding onto this information. You can then go

through some of the steps that we talk about in this guidebook and learn what information is hidden in that data, and make the decisions that your business needs based on this data.

While using your intuition in some cases can still be a good idea, this is not something that we need to rely on as much any longer. Working with big data and a good analysis that can keep up in real-time to this big data, will ensure that you can really take care of the things you need to in business, and can lead you to see some amazing results in the process.

### **Understand Customers**

One way that a lot of companies are going to choose to work with the predictive analysis is in order to understand their customers better. The more that you know about your customers, the easier it is to make the right products for them, enhance some of your current products for them, answer their questions, market to them, and so much more. A company that knows nothing about their customers, or who acts like they barely care about their customer is the company that is going to fail.

This predictive analysis is going to ensure that the company is going to make their customers happy, as long as they really pay attention to what the information is saying, and they are willing to take the right steps to make it happen. Maybe they gather some information from social media that their customers would like more options to choose from, or they wish it were easier to get ahold of the company. The customers may make complaints about the time it takes to ship the items, or that they feel that the item is not really much different than what the competitor has.

While every company would like to think they offer the best product, and it is hard to hear negatives about your product and the hard work that you are putting in, this is a part of life. Being able to improve your business means that you need to listen to the complaints that they are presenting to you, and then try to fix them as well as possible. This doesn't mean that every complaint is reasonable, or that you have to try and make everyone happy. But often, when you bring in some of the analysis and look through the data, you will start to notice that there are some common trends that show up, and you may realize that maybe it is time to make some changes.

You can also use this data to learn more about some of the demographics that come with your customer base and create better marketing campaigns for them as well. The more that you can learn about the demographics of your customers, the easier it is to reach them. Knowing their age, their interests,

their jobs, their hobbies where they live, what kind of family situation they have, their gender, etc., makes it easier to send out personalized messages to everyone and can increase your sales.

Doing predictive analysis over your target audience is going to ensure you gather the right kind of data that serves to inform you about your customers. Shooting in the dark, or assuming that you know your customer without any research, is basically throwing money down the drain. You may be surprised after doing your research that your target demographics are not what you think, or that there may be another kind of group you can reach in addition to your current audience, to increase your profits overall.

One thing you should note is some of the outliers in your data. Sometimes, the outliers are not going to be that important. They will include a few people who want some suggestions that just aren't possible for your business and how it is set up, or they may include those few customers who had a bad experience with your product. If you look at the outliers and see that it is just a few people, and it seems that by making those people happy, it is going to cost too much or alienate the other customers the majority, then it is fine to ignore them.

But it is still important to look at these outliers. Sometimes, they can show us something new, and alert us to a new trend or a new idea that could help put us above the competition. Perhaps you see that most of your demographics fit in the 25-30-year-old range, but then you look at the outliers and find that there is a significant, albeit smaller group, that fits into another range. If these outliers are all in the same age group, then this could be a potential sign of a new demographic group for you, and it may be worth your time to start marketing to them.

This could be the sign idea with a new product idea that you have. Maybe once you see that you have a different demographic, you decide to do some more research and find out which product, in particular, is going to really help them, and may draw in more people from that demographic. Maybe you will start to change up some of your marketing to include ads for this different demographic as well.

Often, this new area is going to be hidden if you just ignore your information and all of that data you are taking in. It is not going to be near large enough to push out your main demographics or what you are currently doing right now, but it is still large enough to show you some potential if you see it. And because you did the predictive analysis, you may be able to reach this

demographic before some of the other people in your industry do, allowing you to make the most money here and see the best results.

Improve the performance of the business

Many times businesses will try to use the predictive analysis in order to improve their overall performance. They want to not only provide some better customer service, but they want to avoid times when their machines are down and not working, cut out costs of production, and not overwork the employees they have. If you gather up the right kind of data and learn how to use predictive analysis on it, it becomes infinitely easier to figure out how to improve all of these areas, making your business save money, reduce waste, and more efficient in the long term.

The first benefit here is that predictive analysis is going to help the business to save money. When they are able to use the analysis to figure out where things are too expensive, or how to become more efficient, that will automatically cut down on a lot of the costs that they are dealing with. When they can prevent things from going to waste as often when they can prevent machines going down all of the time, and when they get a chance to see what parts of the process are necessary and what parts are not, it is much easier to cut down on costs, and improve the bottom line.

Another benefit is that the business can use predictive analysis to provide better customer service. They will learn what their customers really want, be able to find new and innovative ways to deal with their customers, and can ensure that everything is taken care of in an orderly and timely fashion, something that may have been difficult to do in the past.

We can also look at this in terms of warranties and the amount that is being spent on products that are returned. When a customer gets a defective product, they are going to return it. This costs the business, either in refunding the money or in replacing it, plus all that goes into handling these claims. When predictive analysis is brought on board, looking at ways to reduce the defective products so that they save money from here as well. (This is something that also goes into Six Sigma a bit, so if this is your problem, it is another good topic to explore.)

Predictive analysis can also come into play in terms of monitoring the machines that help make your product and learning when it is likely that a part on the machine is going to wear out or break. When we are able to figure this out ahead of time, it is much easier to schedule the repair or the replacement ahead of time and get that fixed during the night or another



downtime for the company. This keeps the production line moving, rather than losing a lot of time and money trying to get things fixed after the part breaks.

### **Predict Behavior**

The fifth benefit that comes with performing predictive analysis is that the business can use this in order to predict the behavior of their customers, and their competitors. This is going to make it easier for them to create marketing for your customers, find the best products to send out to the market, and make sure that you are going to be able to be ahead of the competition all of the time.

First, you can use this kind of analysis in order to predict the kind of behavior that your customers are going to showcase when it comes to your business. This can help you to see what products to release, hopefully, before the rest of the competition steps in and does the same thing. You want to know how they view your product, how they will use your product, and any other behavior that is going to be important when it comes to your business in particular.

Predicting behavior is not always an easy process to work with for many people. It is going to be hard for a business just to figure out what behaviors they are likely to show. And we are not asking you just to make random predictions without help. Instead, we are looking at you to take all of the data that you have been able to collect over some period of time, and then perform a good predictive analysis on this data, opening up a lot of information and insights that can make this possible.

Your predictions may not be accurate all of the time. There are some people who will stray off the path and not behave in the way that you are predicting all of the time. These will be known as the outliers, and it is something that you will see on a regular basis. The point here is to figure out what the general behavior of your customers is like and then work from there to see how you can best serve them to grow your business.

There are a lot of benefits of using predictive analysis for every business, no matter what kind of industry you are in. This can help you to learn so much about the customers you are working with and the competition you have to go against. With all of that data you are bringing in, it makes sense that you would want to do an analysis of it and figure out what trends and insights are inside for your benefit. Make sure to check out these great benefits that go along with predictive analysis to see why your business would see some

improvements in the way your business performs.

## **Chapter 10: The Top Steps to Help You Create Your Own Predictive Analysis Model**

Now that we have spent a good deal of time talking about what predictive analysis is all about and some of the benefits that come from using this kind of analysis in your own business, it is time for us to take this a bit further and look at creating your own predictive analysis model to use any way that you would like. It is fine to have some of the theories behind this analysis to help you get started, but now it is time for us to actually put it to good use and see what steps we have to take to create and implement this kind of model into our own business.

A successful predictive analytics project is one that we are able to execute step by step. As you start to become more immersed in the details of the project, there are a few milestones and a few steps that you will need to follow along the way. Some of these are going to include:

### **1. Define your business objectives**

How are you supposed to create a model that works, if you don't know what you want it to accomplish or how it is supposed to match up with the business you are running? You need to have some clear goals, and some clear objectives, to help you see the best results in this process.

The project here is going to start out with a business objective that is well-defined. The model needs to be able to address the question that your business has. Clearly stating out the objective that you want to follow at the start of the project, and then keep this in mind the whole time you do the work, can make it easier for you to define what the scope of your project will be. It is also one of the best ways for you to test to see how successful the project is from the beginning, and can help you see if another method is needed, long before you get to the end of the project and waste a lot of time.

### **2. Preparing the Data**

From here, we are going to use some historical data in order to train the model that we want to use. The data here is usually going to be scattered across a variety of sources. You can gather the data anywhere you would like, and often the thing that you want to measure will determine where the data will come from. Some companies rely on the information they get from social media. Some will send out surveys, will collect data from how their customer's shop, or will look to research that has been done on their industry

to provide them with some insights.

Since you are getting the data from so many different sources, it is likely that you will need to take some time to cleanse and prepare the data. Data can come in with duplicate records, which can skew the results that you get and some outliers. And depending on the kind of analysis that you want to do, and the objectives for your business, you need to decide whether it is a good idea to remove or keep these things.

Along the way, you may find that your data has some values missing, it may need some kind of transformation to help it to work, and it could be used to generate some derived attributes that will have more predictive power for your objective. Keep in mind that many times, the quality that you have in your data will determine the quality of the model that you produce, so spend some time preparing the data to get the best results.

### **3. Sampling the Data**

At this point, we need to be able to split up our data so that it is found in two different sets. For this, we need to have a set for training, and then a set for testing out the model. Both of these are important because they allow the model to learn what you want, and then you can test it out to make sure that the model learned the way that it was supposed to.

You are going to build up the model that you are creating using the set of data for training. You can use the set of data for testing to verify the accuracy of the model's output. Doing these steps takes some time, but they are absolutely crucial to the process. If you don't do them, you could run into a big risk of overfitting the model. This means that you train the model with a set of data that is so limited that the model ends up picking out all of the characteristics, both the noise and the signal, that are only true for that set of data, rather than for all of the examples you want.

If you have a model that has overfitted for a specific set of data, it is going to end up performing horribly when you want to run it on some of the other sets of data that come in later, or that you are already storing. A test set of data ensures that the model is performing the way that you want and that you are getting the amount of accuracy that you need in the model.

### **4. Building the Model**

Then we need to move on to the idea of building up our model. Sometimes the business objectives or the data that you have will lend themselves to a

specific model or algorithm that you are able to use. If this is the case for you, go ahead and use that algorithm and see how the model can work. This saves you time and can make the predictive analysis so much easier to work with overall.

Then there are times when the best approach that you should use, even with the given data that you have - or the business objective in place - is not going to be as easy to figure out. This doesn't mean that you give up, though. It simply means that you need to do some research and figure out which method is going to be the best for you. And you may have to try out a few options a bit and see what kinds of predictions you are able to make with them.

As you take the time to explore the data that you have, don't be afraid to run as many different algorithms on it as you can. This allows you to see what happens with each one, and you can compare the outputs that you get here. You can then base your choice of the final model on the results that you got overall from these algorithms when they were put to work. This may not be the case all of the time, but sometimes you will find that the results are going to be shown better when you run an ensemble of models at the same time on the data, and then you can choose the final model that you would like to use by comparing the outputs that you see.

## **5. Deploying the Model that You Want to Use**

After you go through the last step of building up your own model, it is time to actually deploy this model in order to get all of the benefits that come out of it. This is a process that is going to require coordination with some other departments, depending on which model you chose to use, and the overall goal of doing this model.

Your goal here is to build a model that is deployable, whether it is just going to be used in one area of your company or in many different parts of the company. You also want to make sure that you have a good method of presenting your results to business stakeholders in a manner that they are able to understand, and you have to convince others so that they are more likely to adopt the model that you have presented.

Once the model is deployed, you will need to monitor its performance, and you can make new improvements to this model as is needed. It is hard to know whether the model is working the way that you want if no one is monitoring it at all. It is common for a model to decay over time; it is not going to work perfectly forever, so watching how it does and updating it any

time that you need by refreshing it with some new data as it becomes available can ensure that your model will continue to work the way that you want.

If you can follow these steps, and spend some time working with the different algorithms that are available, you will find that this is one of the best ways to put the predictive analysis to work for you. The steps are easy to work with; they just take some time as you gather up the data that you want, clean it and then make sure that you pick out the algorithm that works for your needs.

Even with the right algorithm, you need to make sure that you do training and testing along the way. It is never a good idea to just go with an algorithm and assume it is working and giving out accurate results. This is not worth your time and can lead you to make bad predictions and decisions. Instead, you need to focus on training, then testing the model, and double-checking that things match up the way that you want. And when this happens, you know that you have a predictive model that is going to work for your needs, and that will help you to make good predictions for your company, predictions that are going to be based on facts and data.

## Conclusion

Thank you for making it through to the end of *Predictive Analysis*. Let's hope it was informative and able to provide you with all the tools you need to achieve your goals whatever they may be!

The next step is to start learning the best way that you can create your own predictive analysis for your company. You have already been collecting all of that data for some time now, why not learn what is inside all of that data, and how you can use that to provide better customer service, get your customers to feel better, to beat out the competition, and to pick out new products and services to offer on the market?

This guidebook took some time to explore the idea of predictive analysis, and all that it can do for you. There are a lot of options in regards to how to go about accomplishing this process, but you will find that with predictive analysis, you can be sure that you are approaching the information with the goal of making a prediction when it is all done.

We looked at the many different parts of predictive analysis, including working with data mining, machine learning, statistics, and so much more. All of these can be brought in at one point or another to make it easier to gather the data that you need, to clean it up so that it works with your chosen algorithm, and to ensure that you will actually glean some good information out of the process.

We also spent some time looking at the tips and tricks that you can follow in order to get the most out of any kind of predictive analysis that you do. This process is definitely worth your time, and it is going to help you get so much done and beat out the competition. You want to make sure that you are using the process in the right manner, to ensure you get the best results. The tips and suggestions that we provide in this guidebook will make sure that this happens for you.

When you are ready to work on your own predictive analysis, and you want to finally take all of that information you have been gathering and storing and put it to good use, make sure to check out this guidebook to see exactly how you can work on your own predictive analysis today.

Finally, if you found this book useful in any way, a review on Amazon is always appreciated!

If you enjoyed this book, a review on Amazon would be greatly appreciated because it helps me to create more books that people want.

[Click here to leave a review for the book.](#)

Thanks for your support!



**Here's another book by Richard Hurley**



**[Click here to check out this book!](#)**